

Award Number: W81XWH-12-1-0323

TITLE: Advanced Lung Cancer Screening: An Individualized Molecular  
Nanotechnology Approach

PRINCIPAL INVESTIGATOR: James Herman, Professor

CONTRACTING ORGANIZATION: Johns Hopkins University, The  
Baltimore, MD 21218-2680

REPORT DATE: August 2014

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command  
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;  
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE August 2014		2. REPORT TYPE Annual		3. DATES COVERED 1 Aug 2013 – 31 July 2014	
4. TITLE AND SUBTITLE Advanced Lung Cancer Screening: An Individualized Molecular Nanotechnology Approach				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER W81XWH-12-1-0323	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) James Herman  E-Mail: hermajj@jhmi.edu				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)  Johns Hopkins University, The  Baltimore, Maryland 21218-2680				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited					
13. SUPPLEMENTARY NOTES Lung Cancer Screening, CT Screening, DNA Methylation Detection, Emphysema Score, Lung Airspace Variability Score					
14. ABSTRACT This grant utilizes complimentary approaches to improve the early detection of lung cancer, with each aim having independent goals and thus separate utility. Our goal is to explore whether detection of DNA methylation changes and enhanced CT evaluations will add to the specificity of lung cancer detection. In the first year of this proposal, we developed an improved panel of genes hypermethylated in lung cancer, with extraordinarily high specificity and sensitivity. These novel genes were used to develop sensitivity methylation specific PCR assays suitable for biologic fluid testing (sputum and plasma). We also optimized the processing of biologic samples to accomplish improved retention of DNA suitable for methylation detection. In the past year, we have combined this improved method with the newly developed PCR detection and begun to examine biologic fluids from patients with and without lung cancer. For plasma studies, we have examined 141 Cancer patients with plasma and 44 non-cancer patients. For sputum studies, we have examined 89 cancer patients for whom we collected sputum and 23 non-cancer patients with sputum (all with benign pulmonary nodules). Initial studies demonstrate the ability to detect differences in the presence of DNA methylation in cancers compared to controls. Will complete the analysis with an additional 72 Cancer positive and 32 non cancer subjects in the remainder of our funded work. In addition, we have utilized CT data sets of subjects with and without lung cancer to optimize CT as a potential screening tool beyond simply looking for lung nodules. We have data that suggests that parenchymal HU variability may be a potential marker of lung cancer. In combination with these molecular detection approaches, the parenchymal CT imaging approaches should lead to improved screening methods.					
15. SUBJECT TERMS :none listed					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	33	19b. TELEPHONE NUMBER (include area code)

## Table of Contents

	<u>Page</u>
<b>1. Introduction.....</b>	<b>4</b>
<b>2. Keywords.....</b>	<b>4</b>
<b>3. Overall Project Summary.....</b>	<b>4</b>
<b>4. Key Research Accomplishments.....</b>	<b>8</b>
<b>5. Conclusion.....</b>	<b>8</b>
<b>6. Publications, Abstracts, and Presentations.....</b>	<b>8</b>
<b>7. Inventions, Patents and Licenses.....</b>	<b>8</b>
<b>8. Reportable Outcomes.....</b>	<b>8</b>
<b>9. Other Achievements.....</b>	<b>8</b>
<b>10. References.....</b>	<b>8</b>
<b>11. Appendices.....</b>	<b>8</b>

## INTRODUCTION

This grant utilizes complimentary approaches to improve the early detection of lung cancer, with each aim having independent goals and thus separate utility. Our goal is to explore whether detection of DNA methylation changes and enhanced CT evaluations will add to the specificity of lung cancer detection. This is defined in our aims.

Specific Aim 1: To improve the clinical utility and effectiveness of a nested, gel based DNA methylation assay for sputum and plasma by increasing its sensitivity and specificity through nanotechnology. Hypothesis: Detection of DNA methylation from individuals with cancer can be used to determine lung cancer risk and can be enhanced through discovery of optimal hypermethylated genes and implementation of enhanced detection technologies.

Specific Aim 2: To use an in vitro molecular testing of sputum and serum with DNA methylation rather than simple demographics alone to select the highest risk smokers for an expensive screening modality such as CT scanning. Hypothesis: DNA methylation testing is more specific in selecting those at the highest risk for lung cancer than clinical demographics alone.

Specific Aim 3: To optimize low dose chest CT screening for lung cancer. Hypothesis: Valuable information on the chest CT scan, based on the severity, distribution, and pattern of low attenuation areas (“emphysema”), may be crucial to increasing our insights and effectiveness of determining lung cancer risk, the frequency of follow up scans, reducing false positives, and controlling costs compared to an annual chest CT screening for the sole use to detect lung cancer tumors after they occur.

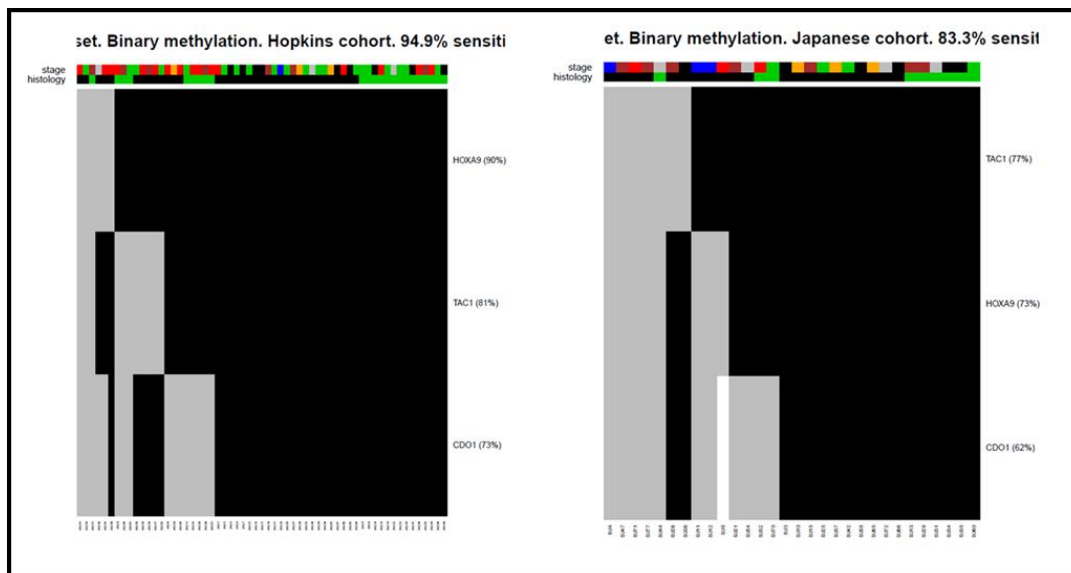
## KEYWORDS

Lung Cancer Screening, CT Screening, DNA Methylation Detection, Emphysema Score, Lung Airspace Variability Score.

## OVERALL PROJECT SUMMARY

For specific aim 1, building upon the progress made in last year’s progress report we have made significant progress on the two sub-aims of this proposal in implementing the developments from last year. Last year’s progress included A) Developing optimal hypermethylated gene panels for detection of tumor DNA from lung cancer and B) Optimize nanotechnology based detection of DNA methylation for increased sensitivity and specificity.

The first efforts were initially focused on the development of an optimal gene panel for detection of lung cancer. After completion of these studies, we have published the results earlier this year<sup>(1)</sup>, with a summary provided here. Hypermethylation of CpG islands is a common and important alteration in the transition from normal to transformed cells. Following previously validated methods for the discovery of cancer-specific hypermethylation changes from NSCLC cell lines, we identified >300 candidate genes. Using the Cancer Genome Atlas (TCGA) and employing extensive filtering to refine our candidate genes for the greatest ability to distinguish tumor from normal, we define a three-gene panel, CDO1, HOXA9, and TAC1, which we subsequently validate in two independent cohorts of primary NSCLC samples. This 3-gene panel is 100% specific, showing no methylation in 75 normal lung tissues from TCGA and 7 normal lung samples from our cohort, and is 83-99% sensitive for NSCLC (shown in last year’s progress report). Our subsequent validation of this in two independent cohorts reveals tumor sensitivity of 95% in a US population from with a lower sensitivity in a cohort from Japan (83%), shown in figure 1). This may reflect the higher incidence of EGFR mutant lung tumors in Asian populations than in the Baltimore region. Our plans for implementation of this panel for detection in plasma and sputum were outlined last year and have progressed well.



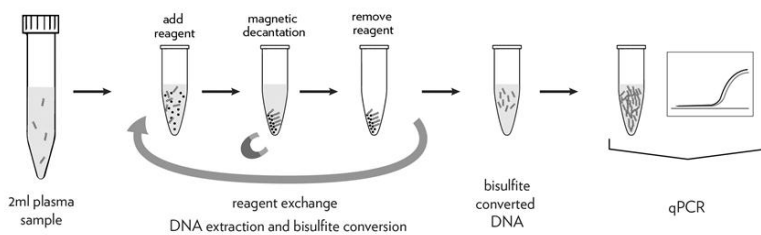
**Figure 1. Methylation of *CD01*, *HOXA9*, and *TAC1* is Highly Sensitive for NSCLC in the Validation Studies.**

Three highly prevalent methylation sites were chosen from data generated within the TCGA studies and real-time MSP analyses assays developed for detection of methylation in lung cancer samples and normal controls. The

tumor results are shown in with high prevalence of methylation in these independent tumor samples, with tumors from Johns Hopkins on the left and from a separate cohort from Japan on the right.

This panel has been further expanded through the identification of additional genes with extremely high methylation frequencies in lung cancer. This panel now includes three additional genes, *HOXA7*, *SOX17* and *ZFP42*, for which real-time MSP analyses assays were also developed to complement the previous 3 gene panel to provide redundant tumor coverage to optimize detection. These new assays were confirmed to specifically detect abnormal methylation using normal lymphocytes and *in vitro* methylated bisulfite converted DNA. We found high specificity to methylation in bisulfite converted DNA and no amplification in unconverted and no template controls. The measured amplification efficiency for all of these genes was 100+/-20%, with assay optimization continuing for *ZFP42*. We are now validating this panel of six genes with clinical samples: We have begun testing gene methylation in normal and cancer patients' sputum, as well as normal lung tissue and lung tumors, after improving the method of DNA processing outlined in aim 2.

**Aim 2:** The use of methylated tumor-specific circulating DNA has shown great promise as a potential cancer biomarker. The relative scarcity of tumor-specific circulating DNA presents a challenge for traditional DNA extraction and processing. We accomplished improvements in DNA processing, with a single tube extraction and processing technique dubbed "methylation on beads" that allows DNA extraction and bisulfite conversion for up to 2 ml of plasma or serum (Outline of approach in figure 2)(2). In comparison to traditional techniques such as phenol chloroform, methylation on beads yields a 1.5 to 5-fold improvement in extraction efficiency. The greatest enhancement in extraction efficiency is seen with small amounts of DNA, precisely matching the need for improved extraction in low DNA content samples such as plasma and serum. A summary of the final results using this approach is provided in figure 3.



**Figure 2. Overview of the Methylation-on-Beads (MOB) Process.** Circulating DNA from up to 2 ml of plasma is extracted and purified via SSBs. The purified DNA is then subject to bisulfite conversion and analyzed via methylation specific PCR (MSP). The entire sample preparation process can be performed in

a single tube and consists of an iterative process of adding reagents, magnetic decantation, and removal of supernatant.

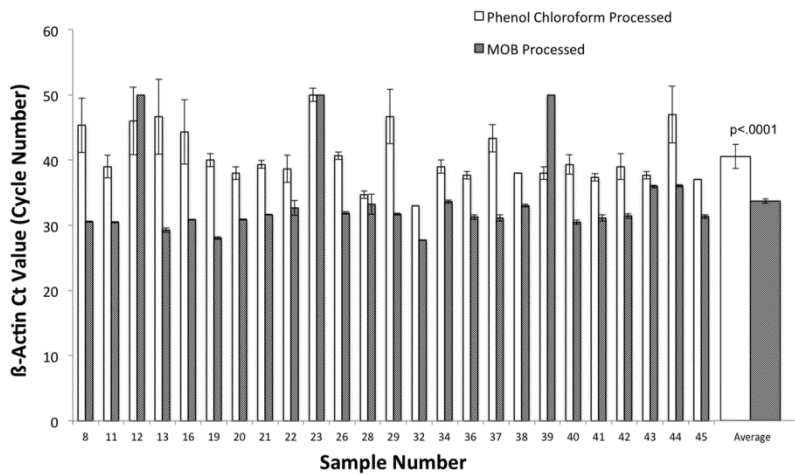


Figure 3  $\beta$ -Actin Ct values for MOB processed vs. Phenol Chloroform extracted and traditionally processed plasma samples from 24 patients diagnosed with lung cancer. The MOB technique demonstrates consistently higher and less variable recovery, as demonstrated by the lower average Ct value (33.8 vs. 40.6 cycles) and Ct standard deviation (0.3 vs. 1.9 cycles), respectively. This improvement in Ct of 6.8 cycles represents a  $2^{6.8}$  or 111 fold increase in amplifiable DNA, on average.

Having developed an optimal panel and improved upon methods for processing the DNA as planned, we have applied these techniques to the plasma and serum of patients with CT detected lung cancer and those with non-cancerous nodules. We have already completed analyses for these 6 genes on the following patients. For plasma studies, we have examined 141 cancer patients with plasma – (stage 1=103; stage 2=13; stage 3=10; stage 4=15). We have also examined 44 non-cancer patients with plasma. This includes age matched medicine clinic population and those found to have benign nodules (the majority of which were granulomas), many of whom were detected through CT screening. We have also examined 89 cancer patients for whom we collected sputum – (stage 1=69; stage 2=5; stage 3=7; stage 4=8), and 23 non-cancer patients with sputum (all with benign pulmonary nodules). For nearly all patients with sputum, there is corresponding plasma that has also been collected and examined. We have 72 cancer positive and 32 non cancer samples left to complete in the remainder of our funded work. We provide a preliminary examination of these results, completed for these 6 genes and with a DNA control (beta actin). All real-time MSP analyses were conducted in triplicates, and will be analyzed according to detection in any PCR reaction as well as the level of detection.

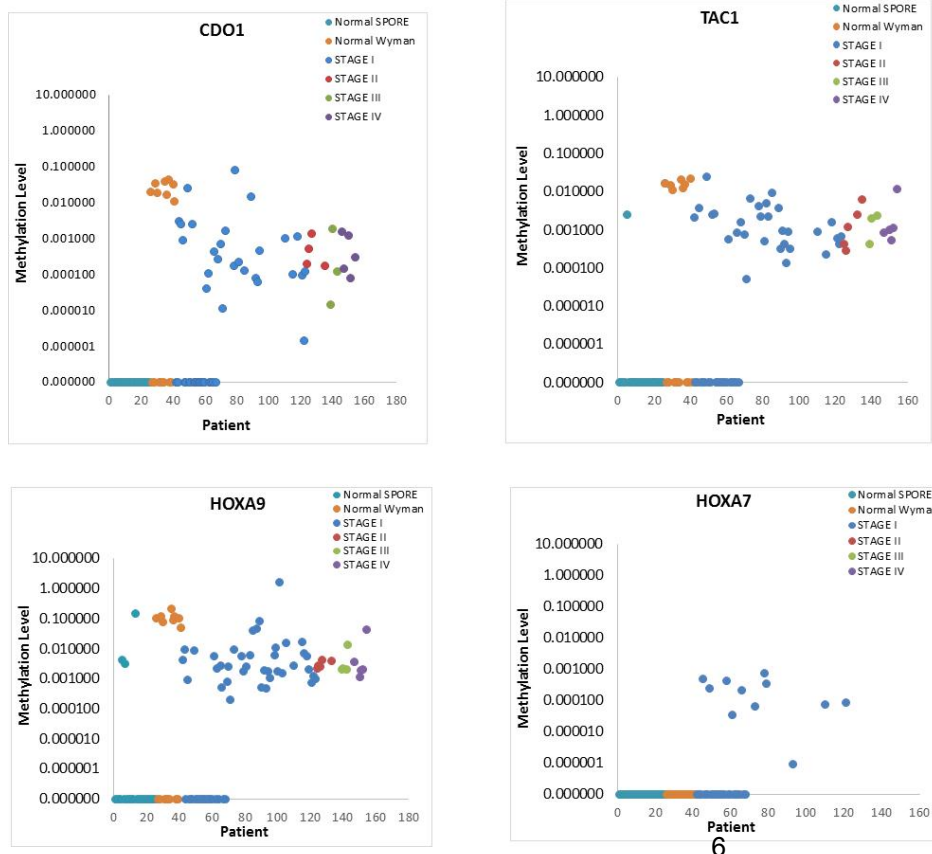


Figure 4. Methylation level (Normalized to beta actin) for 4 genes detected in plasma from patients with Lung cancer and non-cancer controls. Individuals from the two control groups are shown separately, SPORE normal representing CT nodule patients with benign findings, Wyman a medicine clinic control, and the cancer patients shown with different colors according to stage.

For specific aim 3, to date we were able to identify 151 subjects in the SPORE database that had CT scans performed prior to surgeries which were adequate for analysis. We have completed measurement of the extent of computed tomography (CT) in these subjects. Of the group, 127 of the subjects had cancer, and 24 did not. Also, 115 of the subject were current or former smokers with an average of 44 pk/yr history, and 38 of the cancer positive and 4 of the cancer negative were classified as having COPD by spirometry. Because this cohort was heavily biased towards patients with cancer and we needed to include more non-cancer CT studies, we utilized CT data from another study (SCCOR) that included smokers with and without emphysema. In SCCOR, all subjects were without a history of lung cancer and were chest CT negative for any nodules. We also have demographic and pulmonary function data on these subjects. In total we have 127 subjects with a diagnosis of lung cancer and 180 subjects without a diagnosis of lung cancer.

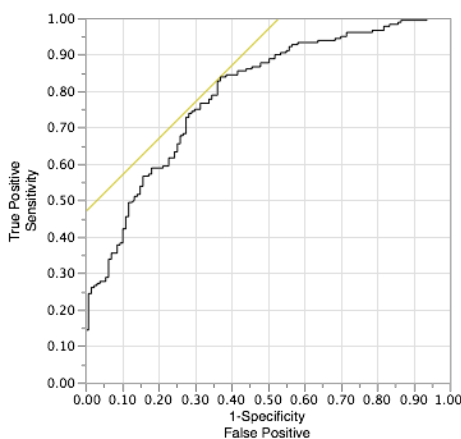
The software can divide the lung into upper, middle, and lower fields on the right and the left for a total of six lung areas. For the subjects, clearly abnormal areas were eliminated from further analysis. For the Ca+ subjects, the final usable number of lung fields were right upper=106, right middle=111, right lower=108, left upper=118, left middle=118, and left lower=116. For the Ca- subjects, we have 103 for each lung field.

The emphysema score was based on the number of voxels with Hounsfield units (HUs) less than -910. The percent emphysema of the lungs ranged from 0.001 to 64.8% among all the subjects with a mean score of  $22.5 \pm 19\%$  (mean  $\pm$  SD). The subjects without cancer had a higher amount of emphysema than those with cancer, which confirmed our results obtained last year with a smaller number of non-cancer CT studies. For the subjects with cancer the mean emphysema score was  $16.8 \pm 16\%$  and for the subjects without cancer it was  $26.5 \pm 20\%$  ( $p < 0.0001$ ). Based on the previous observations that emphysema is an indicator of cancer risk (both related to smoking), then one would have predicted that a higher emphysema score should be associated with a higher cancer risk. However, our data does not support that hypothesis, but rather the opposite was observed. Those with less lung damage (lower emphysema score had a higher risk of cancer). Therefore, emphysema is a poor indicator of cancer risk, and suggests that simple screening for emphysema would not allow for detection of lung cancer.

In contrast, we examined the variability in the voxels throughout the lungs, since there had been evidence in our previous study that this variability, or heterogeneity, was increased in HIV patients with lung cancer (3). To do this we examined the standard deviation around the mean HU level for the lungs of each subject. There was a significant difference in the variability in the mean HU in the two groups. The Ca- group had an average SD of  $118.9 \pm 16$  while the Ca+ group had an average SD of  $134.5 \pm 28.8$  ( $p < 0.0001$ ). Finally, we performed a

multivariate analysis comparing the SD between the two groups controlling for mean HU, the lung volume and the percent emphysema. Controlling for these variables, there was still a significant difference in the SD between the two groups ( $P < 0.0001$ ). In addition, we constructed a Receiver Operating Characteristic curve that was 0.79 (see Figure 5).

Figure 5. ROC curve for use of variability score for the diagnosis of lung cancer. This ROC curve represents a value of 0.79.



## KEY RESEARCH ACCOMPLISHMENTS

- Completion of development of improved Panel of Genes with Cancer Specific Methylation(1) and methods for Optimized processing of biologic samples for methylation analysis(2).
- Implementation of Methylation studies in plasma for 141 Cancer patients and 44 non-cancer patients.
- Studies of emphysema and variability scores completed 127 subjects with a diagnosis of lung cancer and 180 subjects without a diagnosis of lung cancer.

## CONCLUSION

In summary, based on our previous year development of an improved panel of genes hypermethylated in lung cancer, with extraordinarily high specificity and sensitivity, we have examined these novel genes using sensitivity methylation specific PCR assays suitable for biologic fluid testing (sputum and serum) on a cohort of cancer positive and negative samples. In combination with these molecular detection approaches, we have examined the alterations in air space for improving detection of lung cancer and find that variability of air spaces is associated with the presence of lung cancer. The final year's efforts will be the completion of additional patients for molecular detection, publishing the findings from this and the variability score, and comparisons of predictions using the molecular and CT findings.

## PUBLICATIONS, ABSTRACTS, AND PRESENTATIONS

Wrangle J, Machida EO, Danilova L, Hulbert A, Franco N, Zhang W, Glöckner SC, Tessema M, Van Neste L, Easwaran H, Schuebel KE, Licchesi J, Hooker CM, Ahuja N, Amano J, Belinsky SA, Baylin SB, Herman JG, Brock MV. Clin Cancer Res. 2014 Apr 1;20(7):1856-64. PMCID: 4019442

## INVENTIONS, PATENTS AND LICENSES

Nothing to report

## REPORTABLE OUTCOMES

Nothing to report

## OTHER ACHIEVEMENTS

Nothing to report

## REFERENCES

1. J. Wrangle et al., Functional identification of cancer-specific methylation of CDO1, HOXA9, and TAC1 for the diagnosis of lung cancer. Clin Cancer Res 20, 1856 (Apr 1, 2014).
2. B. Keeley et al., Extraction and processing of circulating DNA from large sample volumes using methylation on beads for the detection of rare epigenetic events. Clin Chim Acta 425, 169 (Oct 21, 2013).
3. A. Hulbert et al., Prospective CT screening for lung cancer in a high-risk population: HIV-positive smokers. J Thorac Oncol 9, 752 (Jun, 2014).

## APPENDICES

Included references



# Prospective CT Screening for Lung Cancer in a High-Risk Population

## HIV-Positive Smokers

Alicia Hulbert, MD\*, Craig M. Hooker, MPH†, Jeanne C. Keruly, MS‡, Travis Brown, MS\*, Karen Horton, MD§, Elliott Fishman, MD§, Kristen Rodgers, BS†, Beverly Lee, MS†, Celis Sam, CRNP†, Salina Tsai, MD§, Elizabeth Weihe, MD§, Genevieve Pridham, BS†, Brad Drummond, MD||, Christian Merlo, MD||, Maria Geronimo, RN†, Michelle Porter, RN†, Solange Cox, MD†, Dan Li, BS\*, Marian Harline, RN†, Mario Teran, MD†, John Wrangle, MD\*, Beatrice Mudge, RT§, Gregory Taylor, MD¶, Gregory D. Kirk, MD‡, James G. Herman, MD\*, Richard D. Moore, MD‡, Robert H. Brown, MD#, and Malcolm V. Brock, MD\*†

**Background:** Epidemiological evidence suggests that HIV-infected individuals are at increased risk of lung cancer, but no data exist because large computed tomography (CT) screening trials routinely exclude HIV-infected participants.

**Methods:** From 2006 to 2013, we conducted the world's first lung cancer screening trial of 224 HIV-infected current/former smokers to assess the CT detection rates of lung cancer. We also used 130 HIV-infected patients with known lung cancer to determine radiographic markers of lung cancer risk using multivariate analysis.

**Results:** Median age was 48 years with 34 pack-years smoked. During 678 person-years, one lung cancer was found on incident screening. Besides this lung cancer case, 18 deaths (8%) occurred, but none were cancer related. There were no interim diagnoses of lung or extrapulmonary cancers. None of the pulmonary nodules detected in 48 participants at baseline were diagnosed as cancer by study end. The heterogeneity of emphysema across the entire lung as measured by CT densitometry was significantly higher in

HIV-infected subjects with lung cancer compared with the heterogeneity of emphysema in those without HIV ( $p \leq 0.01$ ). On multivariate regression analysis, increased age, higher smoking pack-years, low CD4 nadir, and increased heterogeneity of emphysema on quantitative CT imaging were all significantly associated with lung cancer.

**Conclusions:** Despite a high rate of active smoking among HIV-infected participants, only one lung cancer was detected in 678 patient-years. This was probably because of the young age of participants suggesting that CT screening of high-risk populations should strongly consider advanced age as a critical inclusion criterion. Future screening trials in urban American must also incorporate robust measures to ensure HIV patient compliance, adherence, and smoking cessation.

**Key Words:** HIV, Lung cancer, Computed tomography screening, Lung cancer screening, High-risk populations.

(*J Thorac Oncol.* 2014;9: 752–759)

Departments of \*Oncology, †Surgery, ‡Medicine, Infectious Diseases, §Radiology, ||Pulmonology, ¶Medicine, and #Anesthesiology and Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland.

Disclosure: The authors declare no conflict of interest.

The first two authors are co-first authors and the last four authors are co-corresponding authors.

Trial Registration: Screening for Lung Cancer in the HIV Patient. NCT01748136 <http://clinicaltrials.gov/show/NCT01748136>.

Address for correspondence: Malcolm V. Brock, MD, Division of Thoracic Surgery, Department of Surgery, Johns Hopkins Medical Institutions, 600 North Wolfe Street, 240 Blalock Building, Baltimore, MD 21117. E-mail: mabrock@jhmi.edu; Robert H. Brown, MD, Department of Anesthesiology and Critical Care Medicine, Johns Hopkins University School of Medicine, 600 North Wolfe Street, JHOC B165 Anesthesiology & CCM, Baltimore, MD 21287. Email: rbrown@jhmi.edu; James G. Herman, MD, Department of Oncology, Johns Hopkins University School of Medicine, Sidney Kimmel Comprehensive Cancer Center, 1650 Orleans Street, The Bunting-Blaustein Cancer Research Building, Room 543, Baltimore, MD 21231. E-mail: hermanji@jhmi.edu; or Richard D. Moore, MD, Department of Medicine, Johns Hopkins University School of Medicine, 600 North Wolfe Street, 1830 Building, Suite 8059, Baltimore, MD 21287. E-mail: rdmoore@jhmi.edu

Copyright © 2014 by the International Association for the Study of Lung Cancer  
ISSN: 1556-0864/14/0906-0752

HIV-infected smokers are reported to have a higher relative risk of developing lung cancer compared with that in the general population, and lung cancer has emerged as the most common and fatal non-AIDS-associated malignancy in most western nations.<sup>1–10</sup> After controlling for cigarette smoking, the best epidemiological estimates are that HIV infection increases lung cancer risk by 2.5-fold.<sup>11–14</sup> Because lung cancer increases markedly with age and duration of smoking, lung cancer may become more common and account for even more deaths as HIV-infected patients live longer with highly active antiretroviral therapy (ART).

The high case fatality rate in HIV-associated lung cancer has been shown not to be attributable to HIV-related causes, but instead, is primarily attributed to an advanced stage of lung cancer presentation in HIV patients.<sup>15,16</sup> Late lung cancer diagnoses occur even in HIV specialty clinics where frequent chest radiographs evaluating opportunistic pulmonary infections fail to detect lung cancer early.<sup>15,17</sup> In fact, approximately 130 HIV-infected lung cancer patients have presented to our institution with more than 80% having late-stage disease.<sup>15</sup>

Since the 1990s, computed tomography (CT) has been explored as an early detection strategy for lung cancer<sup>18–22</sup> with suggestions that it may allow early-stage diagnosis and definitive treatment.<sup>23,37</sup> The National Lung Cancer CT Screening Trial (NLST) reported a 20% reduction in mortality associated with annual CT screening for older, heavy smokers at high risk for lung cancer.<sup>24</sup> Given the current late stage of presentation of HIV-associated lung cancer, CT screening may have profound implications for improving earlier diagnosis of this high-risk group of patients. There are no data, however, to support routine lung cancer screening in HIV-infected smokers because most CT screening studies, including the NLST, excluded their enrollment.

Given that no HIV-infected subjects were enrolled in the NLST, the late-stage presentation of HIV-associated lung cancer, and the epidemiological evidence suggesting this population was at particularly high risk for lung cancer, we hypothesized that annual CT screening in HIV-infected smokers may improve early lung cancer detection. From 2006 to 2013, we initiated a single-armed, prospective, observational study assessing the incidence and stage at diagnosis of lung cancer among HIV-infected smokers undergoing annual CT screening. The primary objective was to determine the prevalence and incidence of lung cancer in HIV-infected smokers. Our secondary objectives were to evaluate the feasibility and adherence to intensive screening in this population, to examine rates of false-positive nodule detection, to determine whether CT screening could change the stage distribution of HIV-associated cancer to that of an early-stage disease, and to determine radiographic markers that may differentiate between HIV-infected smokers with and without lung cancer.

## PATIENTS AND METHODS

### Participants

From January 2006 through May 2013, HIV-infected smokers were recruited and followed from HIV outpatient clinics throughout Baltimore City and from the AIDS Linked to the Intra-Venous Experience cohort at Johns Hopkins.<sup>9</sup> The study was approved by the Johns Hopkins Institutional Review Board, and all subjects provided informed consent. Eligible participants were seropositive for HIV by enzyme-linked immunosorbent assay, had no symptoms of a lung malignancy, aged 25 years or older, and current or former smokers (quit within 15 years) with 20 pack-years of use or more. Exclusion criteria included chest CT examination 18 months before eligibility, pregnancy, history of lung cancer, active respiratory infection, or prior cytotoxic therapy within 6 months. A total of 236 participants were registered. Twelve subjects were excluded because of a CT scan within 18 months of registration leaving 224 participants. Forty-nine participants from the AIDS Linked to the Intra-Venous Experience study were registered from 2010 to 2011 and consented to undergo baseline and end of study imaging only. All enrolled subjects were unselected, and no preference was given toward recruiting “healthier” smokers.

Patient navigators tracked all study appointments, including contacting subjects before appointments, providing minimal

financial remuneration for attendance at each visit, and coordinating follow-up study visits with routine clinical care.

### Screening

At baseline, smoking habits, general health, occupational, and contact data were recorded, and portable spirometry was performed. Forced vital capacity and forced expiratory volume in 1 second were measured at each CT screening. Participants were to have a low-dose helical CT scan at baseline (T0) and up to four scans annually (T1–T4). CT screenings were without contrast using a low-dose regimen (120 kVp, 50–200 mA, 1–5 mm axial reconstruction, 1.1 pitch with collimation of 64×0.6 mm) on a single multidetector scanner (SOMATOM 64; Siemens Medical Solutions, Erlangen, Germany) with daily calibration. Each CT was read independently by two radiologists with interobserver variability ameliorated through joint discussion. Due to previous work in evaluating CT changes in HIV patients, we presumed that CT screening would yield a high incidence of inflammatory nodules and scarring from previous pulmonary infections in HIV-positive patients.<sup>25</sup> Our protocol thus differed from the current, robust protocols of International Early Lung Cancer Action Project (I ELCAP)<sup>26</sup> and the National Comprehensive Cancer Network (NCCN) for CT screening<sup>27</sup> in allowing our radiologists to assess noncalcified pulmonary nodules of 4 to 9 mm diameter as suspicious or nonsuspicious on an individual basis. Repeat low-dose helical CT was recommended at 3 or 6 months for suspicious nodules such as enlarging nodules less than 7 mm diameter or those with other suspicious changes. For nodules 10 mm in diameter or more or enlarging nodules more than 7 mm in diameter, additional diagnostic tests could include CT screening at 3 or 6 months, fluorodeoxyglucose (FDG)-positron emission tomography or Technetium-99m depreotide scintigraphy, or biopsy (percutaneous, bronchoscopic, thoracoscopic, or open biopsy).

### Vital Status

Participants were contacted semiannually enabling updates on health status, contact information, and smoking behavior. The social security numbers of those lost to follow-up were cross referenced with the Social Security Death Index to ascertain vital status. Cause of death was abstracted from the medical record. Data on current CD4 cell count, nadir CD4 count, HIV viral load, and HIV ART were obtained from patients, their health care provider, and from medical records.

### CT Densitometry of Screening Participants

CT scans were analyzed for emphysema using Pulmonary Workstation 2.0 software (Vida Diagnosis, Iowa City, IA). The program determines lung volumes and histogram statistics of all lung pixel attenuation values. Extent of emphysema was estimated by quantifying the percentage of voxels having an attenuation value lower than –910 Hounsfield units (HUs). This threshold was chosen empirically because of the thickness of the CT scans in this study and was validated by analyzing several CT scans over a range of HUs (from –910 to –1040 HU) in 10-HU increments. All lung densitometry measurements were corrected by normalizing to the lung air volume being considered. Of the 224 baseline

scans available, 117 were able to be used for CT densitometry calculations. Two investigators independently performed these analyses with any discrepancies resolved by committee.

## CT Densitometry of HIV-Infected Lung Cancer Patients

From 1989 to 2012 at the Johns Hopkins Hospital, 130 HIV-infected patients were diagnosed with lung cancer. From these patients, 39 had available archived chest CT scan digital data that could be analyzed quantitatively.

## Statistical Analysis

Comparisons of continuous and dichotomous variables between groups were performed with the Student's *t* test (two-tailed) and  $\chi^2$  tests, respectively. Multivariable logistic regression models estimated odds ratios with 95% confidence intervals and were considered significant for *p* values less than 0.05. Statistical analyses were performed with STATA software (Stata Corporation, College Station, TX).

## RESULTS

### Participant Characteristics

We recruited 224 asymptomatic HIV smokers for CT screening (Table 1). At study entry, the median age was 48 years, 90% were black, and 58% had a history of injecting drugs. Most were current smokers (89%) with a median of 34 pack-years smoked. Most had previously received ART. These 224 screened participants were dissimilar demographically to 130 HIV-associated lung cancer patients previously diagnosed at our institution, with the latter being more immunocompromised with higher viral counts and having more obstructive lung disease (Supplementary Table 1, Supplementary Digital Content 1, <http://links.lww.com/JTO/A560>). The age distribution of the screened participants has a bimodal, normal age distribution around a median of 48 years old (Supplementary Figure 1, Supplementary Digital Content 2, <http://links.lww.com/JTO/A561>).

### Adherence to Screening

More than 70% of those eligible patients received both a baseline CT scan and a CT scan in the final year of the study (Fig. 1). The total length of follow-up was 678 patient-years with the median length of follow-up being 3.2 years. After baseline scanning, 44% of eligible patients received a T1 scan, 46% had a T2 scan, 68% received a T3 scan, and fully 71% returned for a final T4 scan (Fig. 1). Participation in each annual screening was hindered by regular changes in residence and frequent alterations in contact information. Of five possible scans for all 224 participants, 18 (8%) received only one scan, 103 (46%) had two scans, 44 (20%) had three scans, 39 (17%) had four scans, and 20 (9%) received all five scans.

### Screening Results

Forty-eight nodules, 32 at baseline and 16 during incident screening, were detected during the study period and followed (Supplementary Table 2, Supplementary Digital

**TABLE 1.** Baseline Characteristics of HIV-Infected Individuals enrolled in the Lung Cancer Detection by CT Screening Study (N = 224)

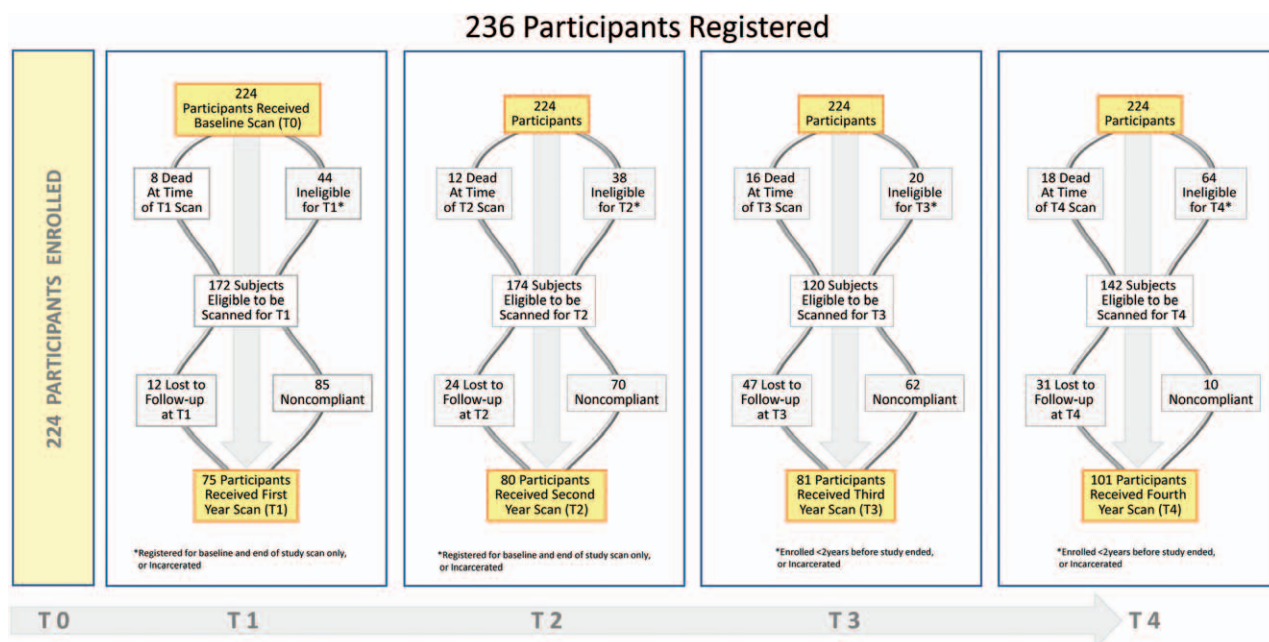
Characteristics	No. of Subjects (%)
Age, median [IQR], yr	48 [44–53]
Sex (M/F)	161/63
Race	
Blacks	201 (90)
Whites	22 (10)
Hispanic or Latino	1 (0.5)
Smoking status	
Former	25 (11)
Current	199 (89)
Never	0 (0)
Pack-years smoked, median [IQR], yr	34 [31–36]
History of marijuana use	90 (40)
History of cocaine use	65 (29)
IVDU ( <i>n</i> = 222)	129 (58)
Hepatitis C ( <i>n</i> = 213)	114 (54)
TB skin-test positive ( <i>n</i> = 210)	44 (21)
STD ( <i>n</i> = 189)	84 (44)
AZT ( <i>n</i> = 209)	143 (68)
CD4 nadir, median [IQR], cells per cubic millimeter ( <i>n</i> = 200)	179 [61–332]
CD4 cell count, median [IQR], cells per cubic millimeter ( <i>n</i> = 187)	400 [217–568]
Viral load <400 cells per cubic millimeter ( <i>n</i> = 207)	123 (59.1)
FEV1, median [IQR], % predicted	85 [70–101]
FVC, median [IQR], % predicted	88 [74–101]
FEV1/FVC, median [IQR],	81 [73–91]
Highest educational level attained ( <i>n</i> = 126)	
Middle school	65 (52)
High school	40 (32)
College degree	21 (17)
Annual income ( <i>n</i> = 87)	
<\$8,000	63 (72)
\$8,000 to \$14,999	12 (14)
\$15,000 to \$24,999	10 (12)
\$25,000 to \$49,999	2 (3)

Some subjects had missing demographic data as noted.

IQR, interquartile range; IVDU, intravenous drug user; TB, tuberculosis; STD, sexually transmitted disease; AZT, azidothymidine; CT, computed tomography; FEV1, forced expiratory volume in 1 second; FEV1/FVC, percentage of the vital capacity which is expired in the first second of maximal expiration.

Content 1, <http://links.lww.com/JTO/A560>). The majority of the 48 nodules were solid; ground-glass consistency represented approximately 30%. Only 25% of nodules were larger than 1 cm in diameter (Supplementary Table 3, Supplementary Digital Content 1, <http://links.lww.com/JTO/A560>). None of these nodules were found to be malignant during subsequent examinations. Of the 48 nodules, 38 were judged not to be suspicious by the radiologists. These included 14 of 38 thought to be caused by chronic inflammation such as from fungal or granulomatous disease, whereas 24 of 38 were thought to have noninflammatory causes such as active infection, scarring





**FIGURE 1.** Flow chart of registered and enrolled HIV smokers. Flow diagram of HIV smokers enrolled in the lung cancer screening study by year of study participation.

from previous infection, or hamartomas. Ten participants with suspicious nodules underwent further CT imaging. Two received a positron emission tomography scan, and only one had a bronchoscopic biopsy. No participant received surgery caused by a false-positive screening.

Although no subject had an interim diagnosis of lung cancer, one non-small-cell carcinoma (NSCLC) of advanced staged (stage 3B) was detected on incident, first-year screening after baseline. The baseline screening CT scan of this patient was at the time not thought to be clinically significant, because the image showed mild hilar adenopathy typical of HIV patients. But, by the time of the first annual T1 screening, there was clear evidence of interval growth in this patient's hilar mass. The patient elected not to have treatment and died 4 months after diagnosis. There were 18 other deaths; all due to causes other than lung cancer. Sixteen of these patients had known CD4 counts at the time of death with 40% (7 of 16) of the patients having a CD4 less than 200 cells per cubic millimeter. Of these seven participants, three died of pneumonia and respiratory failure, two of cancer (tonsillar and pancreatic), and one of encephalitis and renal failure, each."

## Incidental Findings

Of the 224 patients with T0 imaging, 189 participants (84%) had incidental abnormal intrathoracic findings other than suspicious pulmonary nodules. The most often observed intrathoracic abnormalities were emphysematous changes in 69 (37%), pneumonia in 69 (37%), and CT evidence of coronary artery disease in 58 (31%) participants (Supplementary Table 4, Supplementary Digital Content 1, <http://links.lww.com/JTO/A560>). Extrathoracic disease was evident in 40% of patients with the majority being renal and hepatic abnormalities. There was a low prevalence of

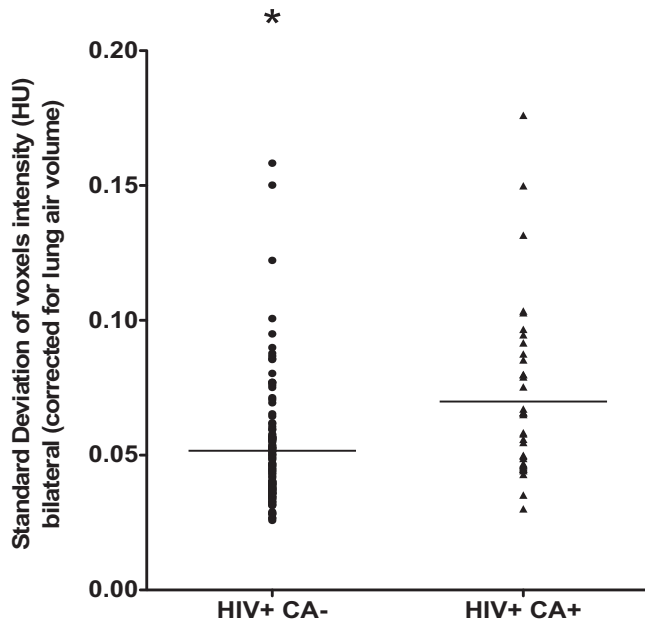
incidental findings that prompted further investigations in the chest and abdomen occurring in 1% and 7%, respectively. Moreover, no extrapulmonary lesions suspicious for cancer were identified.

## CT Densitometry

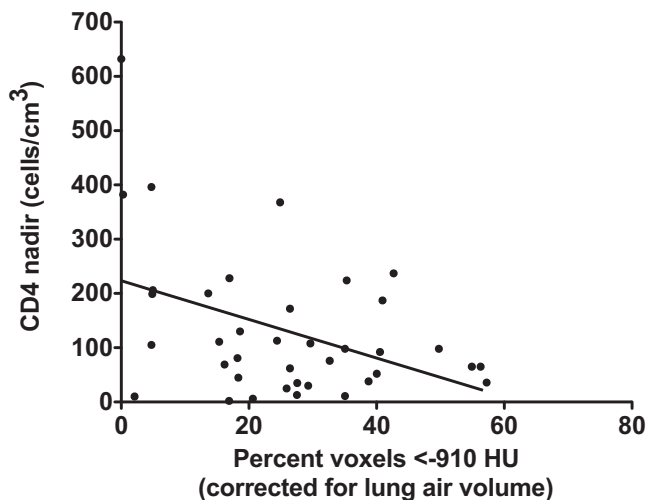
Given the high rate of emphysema detected incidentally and emphysema's high predictive value for lung cancer, we performed CT densitometry analyses of HIV-infected subjects with and without lung cancer. Of the 224 HIV-infected subjects with baseline CT scans, 117 (all without lung cancer) had scans suitable for densitometry. Densitometry analyses of the 117 scans from the CT screening study were compared with 39 scans from HIV-infected patients with known lung cancer diagnosed at our institution. These two groups were dissimilar in age, smoking, the use of azidothymidine, and pulmonary function tests (Supplementary Table 5, Supplementary Digital Content 1, <http://links.lww.com/JTO/A560>).

To assess the degree of heterogeneity of bilateral emphysematous changes in these patients, we measured the variability in voxel intensity across both lung fields in those with and without cancer. The SD of voxel intensity, corrected for lung air volume, was significantly higher in those HIV subjects with lung cancer versus those without lung cancer ( $p = 0.0001$ ; Fig. 2).

Because decreased innate immunity has been associated with emphysema both preclinically and clinically,<sup>28,29</sup> we investigated whether there were differences in the association of nadir CD4 counts and CT densitometry changes in HIV-infected subjects with and without lung cancer. With lower nadir CD4 counts in HIV subjects with lung cancer, there was a significant increase in emphysematous changes by CT densitometric scoring ( $p < 0.001$ ; Fig. 3). This inverse



**FIGURE 2.** Heterogeneity of emphysema of 39 HIV-infected patients with lung cancer and 117 HIV-infected smokers without lung cancer as measured by the variability (SD) in voxel intensities, corrected for lung air volume. The SD was significantly higher in those HIV subjects with lung cancer versus those without lung cancer ( $p = 0.0001$ ).



**FIGURE 3.** Inverse correlation between CD4 counts in cells per cubic millimeter and the percentage of voxels with attenuation less than  $-910$  HU (corrected for lung air volume) in 38 HIV-infected patients with lung cancer. Only one HIV-infected individual with lung cancer had a CD4 count  $>400$  cells per cubic millimeter.

correlation was not observed in HIV subjects without lung cancer ( $p = 0.25$ ). Moreover, because only one HIV patient with lung cancer of 39 patients had a nadir CD4 count more than 400 cells per cubic millimeter, the threshold of nadir CD4 may represent a clinical biomarker to identify HIV smokers at increased risk for lung cancer.

**TABLE 2.** Adjusted Odds Ratio in 140 HIV Individuals (117 HIV-Positive Smokers and 39 HIV-Positive Lung Cancer Patients) of Having Lung Cancer Based on Clinical and Radiographic Characteristics

	Odds Ratio	95% Confidence Interval	<i>p</i> Value
Clinical characteristics			
Increasing age	1.08	1.01–1.15	0.02
Increasing pack-years	1.09	1.04–1.15	$<0.0001$
Decreasing CD4 nadir	1.006	1.002–1.01	0.006
Increased SD/TLV	1.23	1.03–1.47	0.02

Logistic regression model includes subjects' age in years, pack-years cigarette smoking history, CD4 nadir counts (continuous), SD/TLV (SD of voxel intensities corrected by total bilateral lung air volumes), and the percentage of voxels less than  $-910$  Hounsfield units corrected by total lung volume.

Using our 117 HIV-positive patients from our screening cohort and our 39 known HIV-positive lung cancer patients, we performed a multivariate logistic regression analysis assessing clinical and radiographic risk factors associated with lung cancer in HIV-infected patients. Increased age, higher pack-years of cigarette smoking, low CD4 count nadir, and increased heterogeneity of emphysema on CT imaging were all significantly associated with lung cancer in HIV patients (Table 2).

## DISCUSSION

This observational study is the first to evaluate CT screening for lung cancer in HIV smokers. Despite 89% of our cohort being active smokers and the epidemiological evidence of a twofold increase in lung cancer incidence in HIV-infected versus non-HIV-infected individuals,<sup>1–9,30–34</sup> we found only one incident cancer during 678 patient-years using up to five annual CT screenings. Even with the limitation of a small sample size, this is a low rate of lung cancer detection despite our appropriate targeting of a community, epidemiologically, at high risk for lung cancer.<sup>35–41</sup> Although selection bias in recruiting “healthier” HIV smokers is a remote possibility, the more plausible explanation for this low detection rate is the cohort's young median age of 48 years. The normal, Gaussian age distribution shows that this recruitment around a median age of 48 years old most likely reflects the range of ages of HIV-positive patients who sought care at our outpatient clinics and does not suggest selection bias in recruitment. We mistakenly hypothesized that the low immunosurveillance associated with HIV infection would be the most powerful risk factor for lung cancer in our cohort and underestimated the significant contribution of advanced age as a risk factor. In 2003, when our trial was initially designed, the median patient age of all HIV-positive patients in our HIV outpatient clinic was 42.4 years. In 2013, with the widespread use of ART, the median age of all outpatient HIV-positive patients has increased to 52.9 years. Indeed, in most CT screening trials involving non-HIV subjects, 55 years old is the minimum age of eligibility for study participation, and in the original I ELCAP CT screening study published in 1999 by Henschke et al.<sup>36</sup> which showed a 2.7% prevalent lung cancer detection

rate, the 1000 participants had a median age of 67 years. Prospective cohort studies have also shown a strong association between advanced age and increased risk of lung cancer in HIV-infected subjects.<sup>42</sup> The low rate of HIV-infected lung cancer in this trial supports the null hypothesis, and the importance of advanced age to lung cancer incidence, even in communities at high risk for lung cancer. As more CT screening programs increasingly develop algorithms to target high-risk populations of smokers, our negative study suggests that the contribution of advanced age as a significant risk factor should not be ignored.

We found that noncalcified nodules 4 mm in diameter or more were common in smokers who were HIV positive, with the majority of them (38 of 48) interpreted as nonsuspicious by our radiologists due to active infection such as tuberculosis and pneumonia, scarring from previous infection, or granulomatous disease. Only 10 of 48 nodules were thought to be suspicious by the radiologists and all participants with these nodules returned for subsequent imaging. These subsequent images all showed definitively that the lesions were not malignant, sometimes even showing nodule regression.

More than 80% of the screened cohort had additional intrathoracic CT abnormalities, including a third with diffuse coronary artery disease, a finding noted previously.<sup>43</sup> Similar to CT screening trials in non-HIV patients, however, incidental abnormalities requiring further diagnostic work-up were few.<sup>44</sup> Such a high rate of additional intrathoracic findings is an interesting observation as it begs the question whether abnormal CT changes in an HIV-positive and HIV-negative patient require similar follow-up, or whether HIV-positive patients on ART have medication-induced changes to the lung and other organs that are new entities which may raise the false-positive rate and cause potential harm to the patient if aggressively pursued.

But, the most frequently observed intrathoracic abnormalities were emphysematous changes to the lung parenchyma. We endeavored to quantify these emphysematous abnormalities to assess whether they may prove to differentiate in an even higher risk subpopulation of HIV smokers. We did this by comparing 117 participants from our CT screening study with 39 HIV-positive lung cancer patients with CT scans who had previously been diagnosed at Johns Hopkins. There is a suggested association between bullous disease and cannabis usage,<sup>45</sup> but the causal link is not established.<sup>46</sup> In this study, however, there was no difference in the cannabis usage between the 117 participants and the 39 HIV-positive lung cancer patients. Interestingly, there was a significant correlation between decreasing nadir CD4 counts and increasing degrees of CT-determined emphysema in those with lung cancer. Because only one HIV patient with lung cancer had a nadir CD4 count more than 400 cells per cubic millimeter, our data suggest that a certain threshold of nadir CD4 counts for CT screening eligibility may target those HIV smokers at particularly high risk for emphysema-mediated CT densitometry changes. This could also be explained, however, by the possibility that our lung cancer patients with HIV, who were older and smoked more, may

have presented for care later from an immune standpoint, with delayed use of ART. However, an increasingly robust literature suggests the significant dose-response relation of decreasing HIV-induced immunity, often measured by CD4 counts, and the increasing risk of non-AIDS-defining malignancies.<sup>47–49</sup> Finally, our multivariate analysis, between the 39 patients with lung cancer previously diagnosed at Johns Hopkins and those 117 patients without the disease from our screening study, also identified low CD4 nadir as a risk factor for lung cancer, along with increased age, higher pack-years cigarette smoking, and an enhanced heterogeneous pattern of emphysema on CT scanning. Injury and inflammation are known to be pivotal in the nonuniformity of emphysema in the lung,<sup>50</sup> and a dysfunctional immune response in HIV subjects may have accentuated the upper lobe-predominant emphysema observed in HIV subjects.

Despite patient navigators and remuneration for continued participation, this study is limited by few eligible subjects returning for all five scans. Longitudinal engagement in regular HIV care in U.S. urban settings also is a limitation to effective antiretroviral treatment.<sup>51</sup> A recent report of 22,984 adult HIV outpatients receiving care in the United States between 2001 to 2009 indicated that only 20.4% of HIV outpatients were retained as patients on a continual basis without interruption or loss to follow-up.<sup>52</sup> Urban HIV cohorts with a high prevalence of polysubstance abuse are especially vulnerable to poor compliance and follow-up rates.<sup>53,54</sup> Many of our participants returned for a final CT screening at study's end with more than 70% of eligible HIV subjects completing at least a baseline and final CT scan. This suggests that only 14% of our original cohort were truly lost to follow-up, and the majority of participants were merely grossly noncompliant.

Because only one lung cancer was detected, we were unable to investigate the secondary endpoint concerning whether CT screening changes the stage distribution of NSCLC in screened HIV patients versus historic controls. The NLST suggests that stage distribution may change in non-HIV individuals, but the aggressiveness of NSCLC in the HIV-infected patient makes this an open question.

Given the results of this pilot screening study, considerable thought must be given concerning the execution of any large screening study in this high-risk population especially given the many other factors that could make a case against lung cancer screening in such persons; such as “over diagnosis bias,”<sup>55</sup> competing mortality over a course of screening, more aggressive cancer types, faster interval progression of cancers, and the personal anxiety, financial burden, and the morbidity because of the work-up of false-positive tests. At the very minimum, we believe that until the median age of HIV smokers increases, the rate of detection by helical CT of HIV-associated lung cancers will remain low. Advanced age and length of exposure to cigarette smoking are strong risk factors for lung cancer, and most CT screening studies use age older than 55 years as an important eligibility criterion. Our identification of biologic and radiographic markers in HIV smokers to define an even higher subpopulation of high-risk individuals may allow algorithms to determine lung cancer risk more effectively, individualize the frequency of



subsequent scans, reduce false positives, and limit the costs of future lung cancer screening trials. Given the high rate of active smokers in an HIV community and the epidemiological data, as the median age of HIV-infected individuals surpasses 55 years in the United States, perhaps a far larger study enrolling older HIV-positive smokers may answer some of the initial questions we raised here. However, for such a study to be feasible in an urban American HIV cohort plagued by polysubstance abuse, considerable measures to ensure patient compliance, adherence, and smoking cessation must also ensue.

## ACKNOWLEDGMENTS

The authors thank Catherine DeAngelis, MD, MPH, for comments on study design; Allison Brooker and Genevieve Pridham, for technical assistance; Nimisha Thuluvath, Sue Lee, and Eli Luong, for patient accrual and follow-up; and Mary Missouri, for administrative guidance at the Johns Hopkins General Clinical Research Clinic.

Supported in part by the NIH Grants 5K23CA117820-01, CA058184, P30-AI42855, 5M01 RR002719-21, R01DA04334, and R01 HL090483.

## REFERENCES

- Bower M, Powles T, Nelson M, et al. HIV-related lung cancer in the era of highly active antiretroviral therapy. *AIDS* 2003;17:371–375.
- Clifford GM, Polesel J, Rickenbach M, et al; Swiss HIV Cohort. Cancer risk in the Swiss HIV Cohort Study: associations with immunodeficiency, smoking, and highly active antiretroviral therapy. *J Natl Cancer Inst* 2005;97:425–432.
- Frisch M, Biggar RJ, Engels EA, Goedert JJ; AIDS-Cancer Match Registry Study Group. Association of cancer with AIDS-related immunosuppression in adults. *JAMA* 2001;285:1736–1745.
- Parker MS, Leveno DM, Campbell TJ, Worrell JA, Carozza SE. AIDS-related bronchogenic carcinoma: fact or fiction? *Chest* 1998;113:154–161.
- Serraino D, Pezzotti P, Dorrucchi M, Alliegro MB, Sinicco A, Rezza G. Cancer incidence in a cohort of human immunodeficiency virus seroconverters. HIV Italian Seroconversion Study Group. *Cancer* 1997;79:1004–1008.
- Engels EA, Brock MV, Chen J, Hooker CM, Gillison M, Moore RD. Elevated incidence of lung cancer among HIV-infected individuals. *J Clin Oncol* 2006;24:1383–1388.
- Engels EA, Pfeiffer RM, Goedert JJ, et al; HIV/AIDS Cancer Match Study. Trends in cancer risk among people with AIDS in the United States 1980–2002. *AIDS* 2006;20:1645–1654.
- Patel P, Hanson DL, Sullivan PS, et al; Adult and Adolescent Spectrum of Disease Project and HIV Outpatient Study Investigators. Incidence of types of cancer among HIV-infected persons compared with the general population in the United States, 1992–2003. *Ann Intern Med* 2008;148:728–736.
- Kirk GD, Merlo C, O' Driscoll P, et al. HIV infection is associated with an increased risk for lung cancer, independent of smoking. *Clin Infect Dis* 2007;45:103–110.
- Seaberg EC, Wiley D, Martínez-Maza O, et al; Multicenter AIDS Cohort Study (MACS). Cancer incidence in the multicenter AIDS Cohort Study before and during the HAART era: 1984 to 2007. *Cancer* 2010;116:5507–5516.
- Cadranel J, Garfield D, Lavalé A, Wislez M, Milleron B, Mayaud C. Lung cancer in HIV infected patients: facts, questions and challenges. *Thorax* 2006;61:1000–1008.
- Engels EA, Brock MV, Chen J, Hooker CM, Gillison M, Moore RD. Elevated incidence of lung cancer among HIV-infected individuals. *J Clin Oncol* 2006;24:1383–1388.
- Shiels MS, Cole SR, Mehta SH, Kirk GD. Lung cancer incidence and mortality among HIV-infected and HIV-uninfected injection drug users. *J Acquir Immune Defic Syndr* 2010;55:510–515.
- Kirk GD, Merlo CA; Lung HIV Study. HIV infection in the etiology of lung cancer: confounding, causality, and consequences. *Proc Am Thorac Soc* 2011;8:326–332.
- Brock MV, Hooker CM, Engels EA, et al. Delayed diagnosis and elevated mortality in an urban population with HIV and lung cancer: implications for patient care. *J Acquir Immune Defic Syndr* 2006;43:47–55.
- Hooker CM, Meguid RA, Hulbert A, et al. Human immunodeficiency virus infection as a prognostic factor in surgical patients with non-small cell lung cancer. *Ann Thorac Surg* 2012;93:405–412.
- Singh A, Misra V, Thimmulappa RK, et al. Dysfunctional KEAP1-NRF2 interaction in non-small-cell lung cancer. *PLoS Med* 2006;3:e420.
- Bazot M, Cadranel J, Khalil A, et al. Computed tomographic diagnosis of bronchogenic carcinoma in HIV-infected patients. *Lung Cancer* 2000;28:203–209.
- Friedman PJ. Imaging studies in emphysema. *Proc Am Thorac Soc* 2008;5:494–500.
- Brock MV, Gou M, Akiyama Y, et al. Prognostic importance of promoter hypermethylation of multiple genes in esophageal adenocarcinoma. *Clin Cancer Res* 2003;9:2912–2919.
- deTorres JP, Bastarrika G, Wisnivesky JP, et al. Assessing the relationship between lung cancer risk and emphysema detected on low-dose CT of the chest. *Chest* 2007;132:1932–1938.
- Maldonado F, Bartholmai BJ, Swensen SJ, Midthun DE, Decker PA, Jett JR. Are airflow obstruction and radiographic evidence of emphysema risk factors for lung cancer? A nested case-control study using quantitative emphysema analysis. *Chest* 2010;138:1295–1302.
- International Early Lung Cancer Action Program Investigators<sup>1</sup>, Henschke CI, Yankelevitz DF, Libby DM, Pasmantier MW, Smith JP, Miettinen OS. Survival of patients with stage I lung cancer detected on CT screening. *N Engl J Med* 2006;355:1763–1771.
- National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;365:395–409.
- Fishman JE, Schwartz DS, Sais GJ, Flores MR, Sridhar KS. Bronchogenic carcinoma in HIV-positive patients: findings on chest radiographs and CT scans. *AJR Am J Roentgenol* 1995;164:57–61.
- Henschke CI, Yankelevitz DF, McCauley DI, Libby DM, Pasmantier MW, Smith JP. Guidelines for the use of spiral computed tomography in screening for lung cancer. *Eur Respir J Suppl* 2003;39:45s–51s.
- NCCN Clinical Practice Guidelines in Oncology. Lung Cancer Screening. Version 1.2014. Available at: Available at [http://www.nccn.org/professionals/physician\\_gls/pdf/lung\\_screening.pdf](http://www.nccn.org/professionals/physician_gls/pdf/lung_screening.pdf). Published June 12, 2013. Accessed June 27, 2013.
- Zhang X, Shan P, Jiang G, Cohn L, Lee PJ. Toll-like receptor 4 deficiency causes pulmonary emphysema. *J Clin Invest* 2006;116:3050–3059.
- Lee SW, Kim DR, Kim TJ, et al. The association of down-regulated toll-like receptor 4 expression with airflow limitation and emphysema in smokers. *Respir Res* 2012;13:106.
- Allardice GM, Hole DJ, Brewster DH, Boyd J, Goldberg DJ. Incidence of malignant neoplasms among HIV-infected persons in Scotland. *Br J Cancer* 2003;89:505–507.
- Dal Maso L, Polesel J, Serraino D, Franceschi S. Lung cancer in persons with AIDS in Italy, 1985–1998. *AIDS* 2003;17:2117–2119.
- Grulich AE, Li Y, McDonald A, Correll PK, Law MG, Kaldor JM. Rates of non-AIDS-defining cancers in people with HIV infection before and after AIDS diagnosis. *AIDS* 2002;16:1155–1161.
- Herida M, Mary-Krause M, Kaphan R, et al. Incidence of non-AIDS-defining cancers before and during the highly active antiretroviral therapy era in a cohort of human immunodeficiency virus-infected patients. *J Clin Oncol* 2003;21:3447–3453.
- Seaberg EC, Wiley D, Martínez-Maza O, et al; Multicenter AIDS Cohort Study (MACS). Cancer incidence in the multicenter AIDS Cohort Study before and during the HAART era: 1984 to 2007. *Cancer* 2010;116:5507–5516.
- Swensen SJ, Jett JR, Sloan JA, et al. Screening for lung cancer with low-dose spiral computed tomography. *Am J Respir Crit Care Med* 2002;165:508–513.

36. Henschke CI, McCauley DI, Yankelevitz DF, et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. *Lancet* 1999;354:99–105.
37. Nawa T, Nakagawa T, Kusano S, Kawasaki Y, Sugawara Y, Nakata H. Lung cancer screening using low-dose spiral CT: results of baseline and 1-year follow-up studies. *Chest* 2002;122:15–20.
38. Sobue T, Moriyama N, Kaneko M, et al. Screening for lung cancer with low-dose helical computed tomography: anti-lung cancer association project. *J Clin Oncol* 2002;20:911–920.
39. Sone S, Li F, Yang ZG, et al. Results of three-year mass screening programme for lung cancer using mobile low-dose spiral computed tomography scanner. *Br J Cancer* 2001;84:25–32.
40. Infante M, Cavuto S, Lutman FR, et al; DANTE Study Group. A randomized study of lung cancer screening with spiral computed tomography: three-year results from the DANTE trial. *Am J Respir Crit Care Med* 2009;180:445–453.
41. National Lung Screening Trial Research Team; Aberle DR, Adams AM, Berg CD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;365:395–409.
42. Guiguet M, Boué F, Cadranet J, Lang JM, Rosenthal E, Costagliola D; Clinical Epidemiology Group of the FHDH-ANRS CO4 cohort. Effect of immunodeficiency, HIV viral load, and antiretroviral therapy on the risk of individual malignancies (FHDH-ANRS CO4): a prospective cohort study. *Lancet Oncol* 2009;10:1152–1159.
43. Lai S, Lima JA, Lai H, et al. Human immunodeficiency virus 1 infection, cocaine, and coronary calcification. *Arch Intern Med* 2005;165:690–695.
44. Jacobs PC, Mali WP, Grobbee DE, van der Graaf Y. Prevalence of incidental findings in computed tomographic screening of the chest: a systematic review. *J Comput Assist Tomogr* 2008;32:214–221.
45. Gill A. Bong lung: regular smokers of cannabis show relatively distinctive histologic changes that predispose to pneumothorax. *Am J Surg Pathol* 2005;29:980–982.
46. Tan C, Hatam N, Treasure T. Bullous disease of the lung and cannabis smoking: insufficient evidence for a causative link. *J R Soc Med* 2006;99:77–80.
47. Reekie J, Mocroft A, Engsig F. Relationship between current level of immunodeficiency and non-AIDS defining malignancies. Presentation at 16th Conference on retrovirus and opportunistic infections, 2009, Montreal, Canada, abstract 860a.
48. Monforte Ad, Abrams D, Pradier C, et al; Data Collection on Adverse Events of Anti-HIV Drugs (D:A:D) Study Group. HIV-induced immunodeficiency and mortality from AIDS-defining and non-AIDS-defining malignancies. *AIDS* 2008;22:2143–2153.
49. Marin B, Thiébaud R, Bucher HC, et al. Non-AIDS-defining deaths and immunodeficiency in the era of combination antiretroviral therapy. *AIDS* 2009;23:1743–1753.
50. Tuder RM, Yoshida T, Fijalkowka I, Biswal S, Petrache I. Role of lung maintenance program in the heterogeneity of lung destruction in emphysema. *Proc Am Thorac Soc* 2006;3:673–679.
51. Westergaard RP, Hess T, Astemborski J, Mehta SH, Kirk GD. Longitudinal changes in engagement in care and viral suppression for HIV-infected injection drug users. *AIDS* 2013;27:2559–2566.
52. Fleishman JA, Yehia BR, Moore RD, Korthuis PT, Gebo KA; HIV Research Network. Establishment, retention, and loss to follow-up in outpatient HIV care. *J Acquir Immune Defic Syndr* 2012;60:249–259.
53. Arici C, Ripamonti D, Maggiolo F, et al. Factors associated with the failure of HIV-positive persons to return for scheduled medical visits. *HIV Clin Trials* 2002;3:52–57.
54. Giordano TP, Visnegarwala F, White AC Jr, et al. Patients referred to an urban HIV clinic frequently fail to establish care: factors predicting failure. *AIDS Care* 2005;17:773–783.
55. Patz EF Jr, Pinsky P, Gatsonis C, et al. Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Intern Med* 2014;174:269–274.





## Extraction and processing of circulating DNA from large sample volumes using methylation on beads for the detection of rare epigenetic events

Brian Keeley<sup>a</sup>, Alejandro Stark<sup>b</sup>, Thomas R. Pisanic II<sup>c</sup>, Ruby Kwak<sup>d</sup>, Yi Zhang<sup>b,1</sup>, John Wrangle<sup>d</sup>, Stephen Baylin<sup>d</sup>, James Herman<sup>d</sup>, Nita Ahuja<sup>d</sup>, Malcolm V. Brock<sup>d</sup>, Tza-Huei Wang<sup>a,b,c,\*</sup>

<sup>a</sup> Department of Mechanical Engineering, Johns Hopkins University, Baltimore, MD 21218, United States

<sup>b</sup> Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21218, United States

<sup>c</sup> Johns Hopkins Institute for NanoBioTechnology, Baltimore, MD 21218, United States

<sup>d</sup> Cancer Biology Program, The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD 21287, United States

### ARTICLE INFO

#### Article history:

Received 10 June 2013

Received in revised form 16 July 2013

Accepted 17 July 2013

Available online 1 August 2013

#### Keywords:

Circulating DNA

Cancer epigenetics

Blood-processing protocols

Quantification

DNA methylation

Methylation specific PCR

### ABSTRACT

The use of methylated tumor-specific circulating DNA has shown great promise as a potential cancer biomarker. Nonetheless, the relative scarcity of tumor-specific circulating DNA presents a challenge for traditional DNA extraction and processing techniques. Here we demonstrate a single tube extraction and processing technique dubbed “methylation on beads” that allows for DNA extraction and bisulfite conversion for up to 2 ml of plasma or serum. In comparison to traditional techniques including phenol chloroform and alcohol extraction, methylation on beads yields a 1.5- to 5-fold improvement in extraction efficiency. The technique results in far less carryover of PCR inhibitors yielding analytical sensitivity improvements of over 25-fold. The combination of improved recovery and sensitivity make possible the detection of rare epigenetic events and the development of high sensitivity epigenetic diagnostic assays.

© 2013 Elsevier B.V. All rights reserved.

### 1. Introduction

The presence of extracellular nucleic acids in the blood of healthy and diseased individuals was initially observed over 70 years ago [1]. While the particular mechanisms for release of DNA into the bloodstream under normal and pathological conditions have yet to be resolved, many paths have been hypothesized [2–4]. The phenomenon of circulating nucleic acids (CNA) has garnered particular interest as of late as both the amount of CNA and their specific characteristics have been shown to correlate with various disease states as well as tissue trauma [3,5,6]. Tumor specific circulating DNA has shown particular promise as a potential biomarker and has been detected and correlated with numerous cancer types including: lung, pancreatic, liver, prostate, and colorectal [2,3,7]. The analysis of circulating DNA may thus serve as a minimally invasive mode of diagnosis, prognosis and monitoring of cancer [5,8].

DNA derived from cancerous tissue often contains abnormal genetic and/or epigenetic modifications [9]. Epigenetic modifications include heritable changes that occur within cells that do not result in alterations

to the primary DNA sequence. Perhaps the most well known form of epigenetic modification, DNA methylation, has been found to play a key role in cancer initiation and progression, often through loss of expression of key tumor suppressor genes. Consequently, DNA methylation remains a potential marker for applications in cancer detection, diagnosis and prognosis [10].

The use of methylated tumor-specific circulating DNA has shown great promise as a cancer biomarker [11], but its use and reliability are often severely hampered by a number of issues, most notably its relative scarcity. While circulating DNA is found throughout the bloodstream, only a small fraction is likely to come from diseased or cancerous tissue. Furthermore, the even rarer population of cancer-specific genetically or epigenetically modified circulating DNA often places inordinate pressure on current DNA extraction and processing techniques [12]. This extreme rarity, coupled with high-loss processing techniques, results in both lower analytical and clinical sensitivity for diagnostic and prognostic tests. There is thus a clear need for improved techniques that allow for more efficacious extraction, processing and detection of rare circulating methylated DNA.

Traditionally, the methylation status of circulating DNA is determined by extracting the DNA from serum or plasma via phenol chloroform and ethanol precipitation (PC), bisulfite treatment of the extracted DNA, followed last by methylation-specific PCR (MSP) or quantitative MSP (qMSP) [13]. This process typically requires many labor-intensive steps as well as transfer between numerous reaction vessels, thus

\* Corresponding author at: Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21218, United States. Tel.: +1 410 516 7086; fax: +1 410 516 7254.

E-mail address: [thwang@jhu.edu](mailto:thwang@jhu.edu) (T.-H. Wang).

<sup>1</sup> Current Affiliation: The Institute of Bioengineering and Nanotechnology, Agency of Science, Technology and Research, Singapore 138669.

resulting in sample loss, long assay times, increased contamination, high rates of operator error and variable data and success rates. Furthermore, traditional DNA extraction methods often retain the PCR inhibitors found in blood that can significantly affect assay reliability [14]. Lastly, most commercial extraction techniques are not amenable to sample volumes larger than 500  $\mu$ l, as may be necessary in order to provide the clinical sensitivity required on rare genetic biomarkers, where concentrations may be as low as a few methylated gene copies per sample volume (or less). In these cases, the ability to extract DNA from larger volumes is particularly advantageous. While lower volume extraction methods can be used in parallel and pooled together, the effluent remains unconcentrated and provides little benefit in downstream reactions unless additional concentration methods are utilized. Likewise, concentration methods that are currently employed result in sample loss and deterioration due to exposure to elevated temperatures and the resulting concentration of PCR inhibitors and nucleases along with the DNA.

In order to address some of these issues, we previously reported the use of a single-tube method for the extraction and analysis of methylated DNA [15]. Here, we introduce an improved technique, an overview of which is shown in Fig. 1, in order to further extend the original paradigm for detection and analysis of exceptionally rare epigenetically-modified circulating DNA in clinical samples. Dubbed “Methylation-on-Beads” (MOB), the process has been significantly amended for use with larger sample volumes (2 ml) and incorporates key improvements in order to retain and process circulating DNA from plasma with 25-fold more analytical sensitivity than current standard techniques, thereby greatly enhancing the clinical sensitivity of circulating DNA-based diagnostics and clearing the way for the detection of rare epigenetic events.

## 2. Material and methods

### 2.1. Genomic DNA samples

CpG methylated HeLa genomic DNA was obtained from New England Biolabs. All samples using genomic DNA were diluted to their respective concentrations using RNase and DNase free water.

### 2.2. Plasma sample preparation

Patient blood samples were obtained from a previous study [8] conducted according to the Declaration of Helsinki and with Institutional Review Board approval. While the original study contained a patient population of 45 (Average Age: 64, 23 Male/22 Female, 39/8/1 White/Black/Asian, 89% current or former smokers), adequate remaining sample volume for this study was obtained from 25 of the 45 patients. All patients had been diagnosed with Stage IV or unresectable metastatic non small cell lung cancer (NSCLC) and previously already received at least one form of chemotherapy, had measurable disease per RECIST

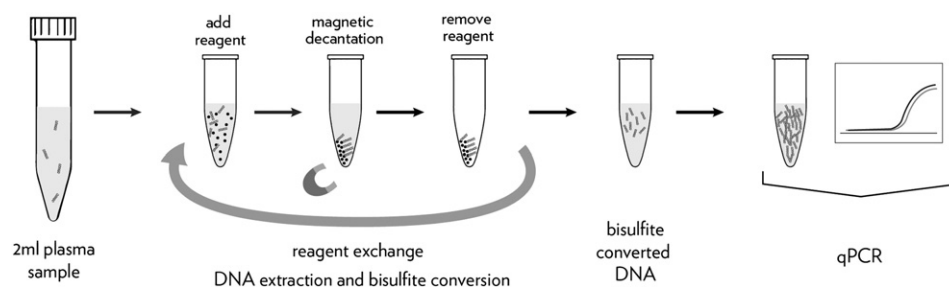
1.0, Eastern Cooperative Oncology Group (ECOG) performance status of 0 to 2, life expectancy > 3 months and adequate liver, renal and bone marrow function. All participants provided written informed consent before participating. Plasma was extracted using standard Ficoll preparation. Briefly, blood samples were immediately placed on ice after draw and, within 60 min, ~10 ml of each blood sample was gently poured onto 3 ml of Ficoll (Sigma-Aldrich) and spun at 1000 g for 10 min. The translucent layer on top was then removed and stored at  $-80^{\circ}\text{C}$  until use.

### 2.3. Large volume methylation on beads process

A 2 ml sample of plasma was digested with the addition of 3 ml of Buffer AL (Qiagen 19075) and 1 ml of Proteinase K (10 mg/ml, Invitrogen) at  $50^{\circ}\text{C}$  for 2–4 h (alternatively, methylated genomic DNA was dissolved in water). Following digestion, 3 ml of 100% IPA and 150  $\mu$ l of SSBs (Promega Magnesil KF - MD1471) were added. The lysate was incubated at room temperature for 10 min to allow for DNA precipitation and binding to the surface of the SSBs. Ten microliter of carrier RNA (1  $\mu\text{g}/\mu\text{l}$ ) was subsequently added to facilitate DNA binding by co-precipitation, and the lysate was again incubated at room temperature for an additional 5 min.

Next, the SSBs containing DNA bound to their surface were isolated and purified from the remaining plasma via magnetic decantation. While the tube remained within the magnetic field, the supernatant was carefully removed without disturbing the isolated SSBs. After discarding the supernatant, the tube was removed from the magnetic holder, and 800  $\mu$ l of Buffer AW1 (Qiagen 19081) was added to the SSBs. The solution was gently vortexed, and transferred by pipette to a 1.5 ml micro-centrifuge tube (for ease of processing). The DNA bound to the SSBs was purified by repeating the steps of SSB isolation within a magnetic field, discarding the supernatant, and removing the tube from the magnetic holder. This process was repeated twice with 500  $\mu$ l of Buffer AW2 (Qiagen 19072). Once the final supernatant was discarded, the remaining supernatant was evaporated off by air-drying within a  $70^{\circ}\text{C}$  heat block for approximately 10 min to remove residual liquid.

In preparation for bisulfite conversion, 45  $\mu$ l of water and 5  $\mu$ l of M-Dilution Buffer (Zymo D5001-2) were added to the SSPs. The solution was incubated at  $37^{\circ}\text{C}$  for 15 min, then 100  $\mu$ l of CT Conversion Reagent (Zymo D5001-1, prepared according to protocol instructions by adding 750  $\mu$ l of water and 210  $\mu$ l of M-Dilution Buffer) was added, and the solution incubated in the dark for 12–14 h. The sample was later cooled down in an ice water bath for 10 min. This was followed by adding 400  $\mu$ l of M-Binding Buffer (Zymo D5001-3) and incubating at room temperature for 10 min. The next step was to add 5  $\mu$ l of Carrier RNA (Qiagen 1017647) and wait for another 5 min at room temperature. After this step, the tube was placed on the magnetic holder and once the SSB were bound to the wall of the tube, the liquid phase was removed and discarded. The particles were then resuspended by adding 400  $\mu$ l of M-Wash Buffer (Zymo D5001-4). The tube was once again



**Fig. 1.** Overview of the methylation-on-beads (MOB) process. Circulating DNA from up to 2 ml of plasma is extracted and purified via SSBs. The purified DNA is then subject to bisulfite conversion and analyzed via methylation specific PCR (MSP). The entire sample preparation process can be performed in a single tube and consists of an iterative process of adding reagents, magnetic decantation, and removal of supernatant.

placed in the magnetic holder and the liquid phase removed. After this wash step, 200  $\mu$ l of M-Desulphonation Buffer (Zymo D5001-5) was added and the sample was incubated at room temperature for 13 min. At the end of this incubation period, an additional 5  $\mu$ l of Carrier RNA were added and the sample was incubated for an additional 3 min. At the end of this step, the tube was placed in the magnetic holder and the liquid phase again removed. Two subsequent wash steps were performed by adding the M-Wash Buffer, placing the tube in the magnetic holder to remove the liquid phase, and repeating. After the liquid was removed for the second time, the tubes were spun down in order to bring the SSB to the bottom as well as release some of the liquid out of them. The tubes were placed on the magnetic holder again to remove this excess liquid. The tubes were then transferred to a hot plate at 90 °C for the ethanol from the wash buffer to evaporate. Once the SSB were dry, 62  $\mu$ l of M-Elution Buffer were added (D5001-6). The SSB were then incubated at 90 °C for 10 min. The tube was placed on the magnetic holder and the liquid transferred to a new tube. Then an additional 50  $\mu$ l were added to the tube containing the SSB and it was incubated at 90 °C for 10 min. The tube was placed on the magnetic holder and the liquid transferred to the same tube containing the 62  $\mu$ l transferred previously. Due to evaporation and the liquid absorbed by the SSB, the final volume yield is ~100  $\mu$ l.

#### 2.4. Phenol chloroform and alcohol extraction

500  $\mu$ l of methylated genomic DNA or processed plasma samples (note that for 2 ml samples all volumes were scaled up by a factor of four) were transferred into microcentrifuge tubes containing 0.5 mL of DNA Extraction Buffer and 100  $\mu$ l of Proteinase K (Sigma Aldrich). The tubes were mixed and incubated at 55 °C overnight. One 2 ml MaXtract gel tube (Qiagen) per sample was spun for 3 min at 15,000 rpm. To each gel tube, an equal volume of Phenol/Chloroform (pH 8.0) and digested sample was added. The gel tubes were then spun for 5 min at 15,000 rpm, separating the phases into the aqueous (above the gel matrix) and the organic (below the gel matrix). Using a pipette, the aqueous (top) layer of each sample was then transferred to a fresh microcentrifuge tube. For each sample, 650  $\mu$ l of 100% EtOH, 200  $\mu$ l of 7.5 M Ammonium Acetate ( $\text{NH}_4\text{Ac}$ ), and 2  $\mu$ l of GlycoBlue was added to each microcentrifuge tube and vortexed. The tubes were then placed in a –20 °C freezer overnight to precipitate for up to 3 days. The microcentrifuge tubes were next spun for 45 min at 15,000 rpm and the precipitate mixture decanted. The resulting pellets were washed with 1 mL of 70% EtOH and spun for 15 min at 15,000 rpm. The supernatant was then discarded, being careful not to dislodge the pellet. Each sample was then air dried in a chemical hood until all the EtOH was evaporated. Lastly, the pellets were resuspended in 100  $\mu$ l of TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0).

#### 2.5. Qiagen commercial kit extraction

For the MOB comparison to commercial kits, the QIAmp Circulating Nucleic Acid kit (Qiagen) was used according to the manufacturer's instructions prior to bisulfite conversion.

#### 2.6. Standard Bisulfite Conversion

DNA recovered by the PC or the Qiagen kit was subject to bisulfite conversion using EZ DNA Methylation™ Kit (Zymo) according to the manufacturer's instructions. The bisulfite conversion buffers used in the single-tube MOB process are the same as those used in the Zymo EZ DNA Methylation Kit. In order to provide a consistent comparison, the DNA extracted using the Qiagen kit and the PC method were bisulfite converted by using the silica matrix spin-columns included in the Zymo kit according to the manufacturer's protocol. The final elution volume was adjusted to 100  $\mu$ l in all cases.

#### 2.7. Methylation-specific PCR and cycle threshold calculation

Two microliter of bisulfite converted DNA target (or equivalent plasmid DNA) was added to 23  $\mu$ l of quantitative PCR reaction mixture. Final reaction conditions were as follows: (10x buffer), 300 nM sense primer, 300 nM anti-sense primer, 100 nM probe, 10 nM fluorescein reference dye (Life Technologies), 200  $\mu$ M dNTPs (Denville Scientific), and a single unit of Platinum Taq® DNA Polymerase (Life Technologies). Thermocycling was controlled as follows: 95 °C for 5 min, 40 cycles of 95 °C for 30 s, 60 °C for 30 s, 72 °C for 30 s within the MyIQ thermocycler (Bio-Rad Laboratories).

The Cycle Threshold (Ct) value is defined as the PCR cycle number that the fluorescence signal surpasses a threshold level. The threshold level was typically calculated by using the computer software provided with the qPCR thermocycler. However, when the computer algorithms were visibly unable to determine the accurate Ct Value, manual and comprehensive alterations were made for the selection of the background fluorescence and threshold level for the comparative samples, as permitted within the software provided by Bio-Rad Laboratories.

#### 2.8. Determination of "positive" clinical samples

Three individual qPCR reactions were performed for each plasma sample. One of the 25 patient plasma samples contained insufficient volume for triplicate measurement, leaving a total of 24 samples for RASSF1A analysis. Due to the relative scarcity of methylated DNA, the patient was considered to be positive for the Ras association domain family 1 isoform A (RASSF1A) gene if at least two of the three qPCR reactions demonstrated DNA amplification with the methylation specific primer set for the RASSF1A gene.

### 3. Results

#### 3.1. Methylation on beads extraction and processing of genomic DNA

We first tested the ability of the MOB process to extract and process genomic DNA dissolved in water as an idealized model system. As DNA concentrations in the plasma and serum of humans typically range from 1 to 1000 ng/ml, depending on the individual and burden of disease [3], we demonstrated the linearity of DNA recovery for the streamlined MOB process using DNA concentrations within this range, as shown in Fig. 2. Here, fully methylated genomic DNA was diluted in 2 ml of water to final concentrations of 1, 10, 100, and 1000 ng/ml. The samples were then subject to the entire MOB process, and qPCR was performed for the  $\beta$ -actin gene as a means of determining the amount of DNA

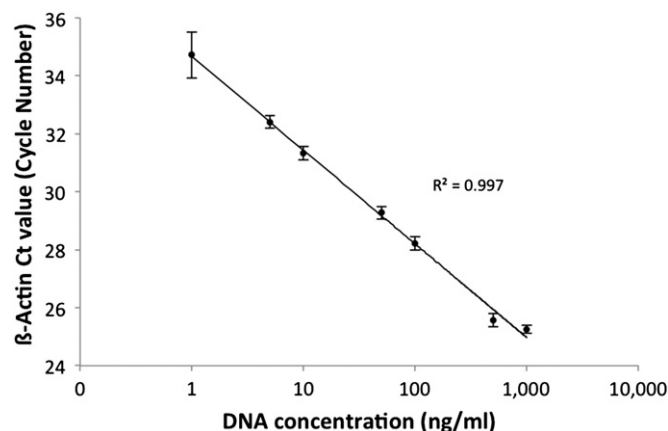


Fig. 2.  $\beta$ -actin cycle threshold (Ct) values of MOB processed DNA versus initial DNA concentration. The Ct value shows an inverse correlation with respect to starting DNA concentrations, thus demonstrating the linearity of the MOB process, from sample preparation to methylation specific PCR, of over 4 orders of magnitude.

recovered [16]. In qPCR, the cycle threshold (Ct) value is the fractional cycle number at which the number of amplified copies reaches a fixed threshold. The Ct value is thus typically used as a means to assess the quantity, either absolute or relative, of DNA present in a sample. From the dilution series, we used the Ct values of each dilution in order to demonstrate that the MOB process provides a linear rate of DNA recovery ( $R^2 > 0.99$ ) and should, in the absence of interfering substances, ostensibly allow for quantification of DNA concentration within this range.

### 3.2. Quantification of methylated genes using MOB

In order to assess the ability of the MOB process to be used for direct quantification of circulating DNA on concentrations as low as 10 copies per reaction volume or fewer, we proceeded to compare the recovery rate and qPCR quantification of genomic DNA against an in-house plasmid standard equivalent to a set number of bisulfite converted adenomatous polyposis coli (APC) gene copies. Thus by comparing the Ct values for the APC gene from MOB processed genomic DNA with standards from known copy numbers of the plasmid DNA, we can compare the [approximate] copy numbers of the input genomic DNA with values back calculated from the plasmid standards. Here we assume 6 pg of diploid genomic DNA per full genome copy. The data is plotted in Fig. 3. Overall, the results between the plasmid APC standard and MOB-processed DNA display a high level of correlation, particularly at lower copy numbers (the region of interest). With a genomic DNA input concentration of 2 ng/ml (corresponding to approximately six gene copies per qPCR reaction), the MOB process yields approximately a 90% recovery in comparison to the expected Ct values based on the plasmid DNA dilution series.

### 3.3. Recovery of DNA via MOB vs. other standard DNA processing techniques

We sought to compare the recovery rate of the improved MOB technique to traditional laboratory techniques (phenol chloroform extraction) and a commercially available kit (Qiagen QIAmp Circulating Nucleic Acid kit). Each extraction and processing technique was tested using methylated genomic DNA diluted into 2 ml of water with final concentrations ranging from 1 ng/ml to 1 µg/ml, corresponding to a total DNA input range from 2 ng to 2 µg. Following bisulfite conversion, the recovered DNA was then quantified using qPCR for  $\beta$ -Actin and the Ct values for each technique compared. Fig. 4 shows the percent recovery using various standard extraction techniques, as compared to the MOB technique. As can be seen from the graph, the MOB technique

showed superior recovery at all concentrations tested particularly at the low input levels (2 ng).

### 3.4. MOB vs. traditional phenol chloroform extraction in human plasma samples

We next compared the performance of the improved MOB technique to traditional PC in the processing of clinical samples. A library of plasma samples from 24 patients diagnosed with Stage IV lung cancer was used for assessment. Using identical starting material, circulating DNA was extracted from 2 ml plasma samples using the improved MOB technique and compared with DNA previously extracted using standard PC from corresponding 500 µl plasma samples. After processing, the samples were quantified, as previously, using qPCR for  $\beta$ -Actin. Fig. 5 shows the results of the two methods in terms of total DNA recovery. The MOB technique shows far lower Ct values (more DNA recovery) than the phenol chloroform method. The difference in average Ct value of 6.8 accounts for a recovery rate of  $\sim 2^{6.8} = 111$ -fold more analytic sensitivity (Paired two-tail t-test,  $p = 1.4 \times 10^{-5}$ ) when using the 2 ml MOB process as compared to the traditional PC technique. Furthermore, the precision of the MOB technique far outperformed phenol chloroform extraction, yielding an average Ct standard deviation of 0.3 cycles, as compared with 1.9 cycles for phenol chloroform.

Lastly, in order to demonstrate the potential for clinical significance of the improved MOB technique, we performed qMSP on the processed DNA from 24 plasma samples for the Ras association domain family 1 isoform A (RASSF1A), a normally unmethylated tumor suppressor gene whose methylation is known to be associated with lung and various other cancers [17,18]. The results of MSP for the RASSF1A gene are shown in Table 1. The samples processed using traditional phenol chloroform extraction and bisulfite conversion methods showed RASSF1A methylation in only 3 of 24 (12.5%) samples, while those samples processed with the MOB technique demonstrated a methylation rate of 42% (10 of 24), a rate that falls within the upper end of the range reported for tissue in lung cancer patients [19]. Improvements such as this may significantly improve both cancer screening, as well as provide increased sensitivity for the monitoring of epigenetic therapies [8].

## 4. Discussion

The relative scarcity of tumor-specific methylated circulating DNA in the bloodstream exerts heavy demands on techniques for the extraction and processing of this DNA for detection and quantification. Likewise, there is a consistent need for new and improved methods that will allow for detection of rare epigenetic events. Here, we sought to introduce and characterize an improved MOB technique for the extraction and processing of methylated DNA. We also compared the MOB technique to other standard methods of DNA processing in terms of extraction efficiency and clinical sensitivity. Overall, the improved MOB technique performed superiorly, compared to both traditional phenol chloroform and alcohol methods, as well as a commonly employed commercial extraction kit.

The streamlined methylation-on-beads (MOB) process utilizes silica superparamagnetic beads (SSBs) as the DNA carrier to integrate DNA extraction and bisulfite conversion into a single platform. SSBs are micro/nanoparticles that are frequently used for solid phase nucleic acid extraction, and commercially available SSB vary in size from 5 nm to 400 µm [20–22]. The silica surface provides a solid substrate for nucleic acid adsorption. The superparamagnetic property allows SSB to be easily manipulated remotely with an external magnetic field, thereby greatly simplifying sample processing.

The general principle of the updated MOB process is illustrated in Fig. 1. In short, the process allows for the ultra-high efficiency extraction of circulating DNA from up to 2 ml of serum or plasma, followed by bisulfite conversion of the cell-free DNA and 20-fold (or more) concentration

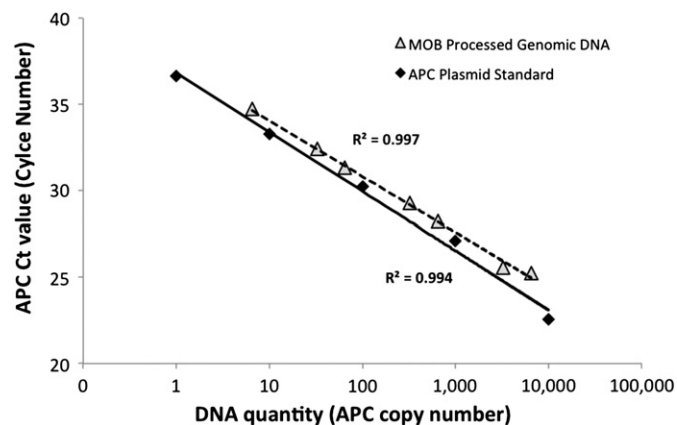
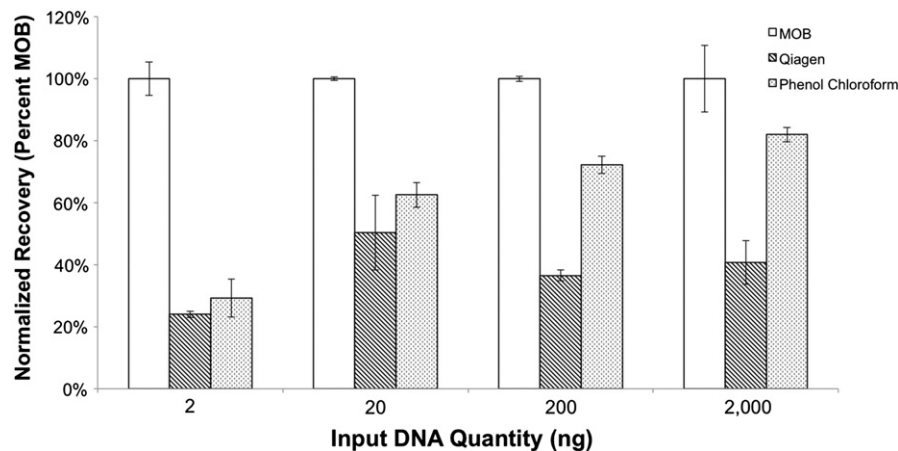


Fig. 3. APC gene Ct values vs. gene copy number for MOB-processed DNA. Ct values of MOB-processed genomic DNA were compared with known copy numbers of APC plasmid DNA. The Ct values show excellent rate of recovery and correlation with respect to the plasmid DNA standard.





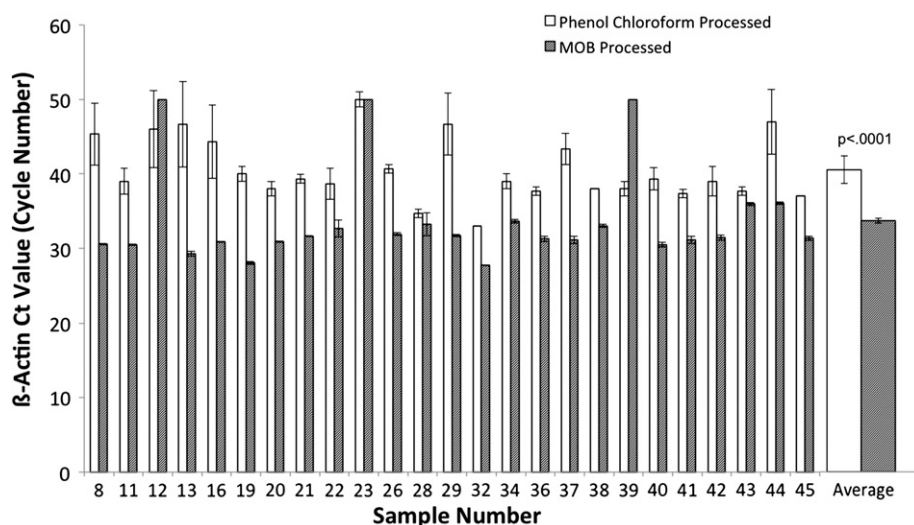
**Fig. 4.** Normalized DNA recovery, as quantified by  $\beta$ -actin qPCR, of MOB compared with traditional phenol chloroform and alcohol (PC) extraction and a Qiagen Extraction Kit. The MOB technique exhibits superior recovery rates at all DNA concentrations tested.

of the sample to reach a final volume of 100  $\mu$ l (or less). The step-by-step procedure follows the simple process of adding the consecutive reagent(s), placing the tube in a magnetic holder to isolate the magnetic particles, removing the supernatant, and repeating the process. This facile method allows for easy implementation within the laboratory setting, and personnel do not need to be extensively trained in order to perform it. In addition, the process utilizes commercially available buffers and reagents in order to increase reproducibility and uniformity across samples and between laboratories.

The entire MOB process takes approximately 16 h to complete, only 4 of which require hands-on benchwork and the remaining 12 are for sample incubation. This processing time is significantly shorter than the widely used PE process that includes lengthy precipitation waiting times and requires at least two days to complete. Furthermore, we have internally verified that the entire MOB process can be reliably performed in as few as five h through incorporation of rapid bisulfite conversion reagents such as those found in the EZ-DNA Methylation-Lightning Kit (Zymo Research; data not shown). This will enable the user to complete the entire sample to analysis process within a single working day. In terms of cost, if the reagents are purchased at medium volume, the price per extraction by the MOB process is approximately ten dollars. The main expense is Proteinase K, which represents 80% of

the cost and would ostensibly be utilized by any DNA extraction method. The overall cost of the MOB process is thus comparable between all the presented methods, including phenol chloroform, which requires a similar amount of proteinase K for the initial digestion and approximately one to three dollars in additional reagents. A phase lock gel tube for extraction such as the Qiagen Maxtract costs approximately 70 cents per unit and is commonly used since it simplifies the process and minimizes contamination, but additionally requires the use of a centrifuge.

Overall, the MOB technique shows a drastically improved recovery rate as compared to traditional PC methods. Even if one accounts for the four-fold increase in starting material (2 ml vs. 500  $\mu$ l), there still remains an over 25-fold increase in signal from the recovered circulating DNA. This can likely be accounted for by at least two advantages of the MOB technique: (1) improved recovery as shown in Fig. 4 and, notably, (2) significantly reduced carryover of PCR inhibitors. The improved DNA recovery of the MOB technique is likely in part due to incorporation of carrier RNA in several key processing steps. The rationale for this inclusion is that while silica exhibits a relatively high affinity for DNA, recovery can sometimes prove problematic in low DNA solutions such as found circulating within the bloodstream. The use of carrier RNA helps facilitate the precipitation of DNA so that it can be more



**Fig. 5.**  $\beta$ -Actin Ct values for MOB processed vs. phenol chloroform extracted and traditionally processed plasma samples from 24 patients diagnosed with metastatic non-small cell lung cancer. The MOB technique demonstrates consistently higher and less variable recovery, as demonstrated by the lower average Ct value (33.8 vs. 40.6 cycles) and Ct standard deviation (0.3 vs. 1.9 cycles), respectively.

**Table 1**

Detection of methylation of the RASSF1A tumor suppressor gene in circulating DNA; MOB vs. traditional phenol chloroform and alcohol extracted and processed plasma samples from 24 patients diagnosed with metastatic non small cell lung cancer. The MOB technique identified 7 more methylation positive samples corresponding to the potential for over 3-fold higher clinical sensitivity than traditional phenol chloroform methods.

Sample #	Positive by phenol chloroform	Positive by MOB technique
8	N	N
11	Y	N
12	N	Y
13	N	Y
16	N	Y
19	N	Y
20	N	N
21	N	Y
22	N	N
26	N	N
28	N	N
29	N	N
32	N	Y
34	N	N
36	N	N
37	Y	Y
38	Y	N
39	N	Y
40	N	N
41	N	N
42	N	Y
43	N	Y
44	N	N
45	N	N
Total positives	3/24 (13%)	10/24 (42%)

readily captured onto the silica surface of the SSB, resulting in significantly higher yields [23].

A particular advantage of SSB-based DNA processing is the ability to extract and concentrate the DNA with little carryover of PCR inhibitors [24]. Silica-coated beads have a specifically high affinity for the adsorption of nucleic acids, thereby providing the ability to readily aspirate away contaminants and inhibitors, particularly when used in a single concentrating step. Thus, by lowering the concentration of PCR inhibition, more DNA can be incorporated into each reaction volume thereby proportionately increasing the detection sensitivity. This is particularly relevant in the case of circulating DNA, as plasma and serum samples are known to contain high levels of numerous PCR inhibitors, thus resulting in false negatives and reduced PCR efficiency [14]. Alternative solutions for DNA concentration have included dehydration using a vacuum manifold, heating, or a combination of both. While these methods do increase the concentration of DNA, they concomitantly increase the concentration of contaminants and PCR inhibitors, a problem that is exacerbated in concentrating larger sample volumes required to detect rare events. This leads to a lower PCR efficiency and higher CT values for a given amount of DNA, which may be misinterpreted as a lower DNA quantity. We independently confirmed this by performing an internal comparison between the concentrated output of ten MOB-processed 200 µl samples and one MOB-processed 2 ml sample following the protocol presented in this manuscript. Our results indicated that the latter showed both significantly better PCR efficiency and consistency across all samples tested (data not shown).

Comparison of the RASSF1A qMSP results between the traditionally-processed and MOB-processed NSCLC samples shows a significantly higher positive rate using the MOB-process. While the MOB-processed RASSF1A positivity rate does indeed fall within upper end of the traditionally reported range for NSCLC tissue samples, implying improved clinical sensitivity [25], the positive predictive value (PPV) could not be confirmed in these studies due to a lack of matching tissue samples to act as a gold standard. Furthermore, the results shown in Table 1 indicate that two of the three samples that were positive for RASSF1A

methylation using the traditional processing techniques were not positive when processed via MOB. Thus, while this study does demonstrate improved clinical sensitivity, further studies, particularly using matching tissue samples, will be required to verify the utility of the MOB process for improved PPV. Likewise, while the genes tested in this study, RASSF1A and APC, are rarely methylated in healthy individuals [18,26], future studies that include healthy samples will be required in order to investigate the effect of the MOB process on negative predictive value (NPV).

Sometimes taken for granted, improved sample processing techniques can dramatically improve assay sensitivity, both analytical, as demonstrated in Fig. 4, and clinical, as demonstrated in Table 1. These improvements may appreciably impact clinical care through reliable detection of epigenetic events, such as methylation, at earlier stages, thus allowing for prophylactic measures to be undertaken in order to avoid cancer initiation and/or progression.

## 5. Conclusions

Methylation-on-beads represents a simple, but efficacious method for the extraction and processing of circulating DNA in preparation for methylation specific PCR. Its numerous advantages include: simplicity, use of commercial off-the-shelf reagents, high DNA retention and little carryover of PCR inhibitors resulting in significantly improved sensitivity for the detection of rare epigenetic events.

## Acknowledgements

The authors greatly appreciate all of Kristen Rogers' work in the preparation of the plasma samples. We also would like to thank the National Institutes of Health (R01CA155305, U54CA151838, R21CA173390) for funding this work.

## References

- [1] Mandel P, Metais P. Les Acides Nucleiques du Plasma Sanguin Chez L'Homme. *C R Seances Soc Biol Fil* 1948;142:241–3.
- [2] Jung K, Fleischhacker M, Rabien A. Cell-free DNA in the blood as a solid tumor biomarker—a critical appraisal of the literature. *Clin Chim Acta* 2010;411:1611–24.
- [3] Schwarzenbach H, Hoon DSB, Pantel K. Cell-free nucleic acids as biomarkers in cancer patients. *Nat Rev Cancer* 2011;11:426–37.
- [4] Jahr S, Hentze H, Englisch S, et al. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res* 2001;61:1659–65.
- [5] Fleischhacker M, Schmidt B. Circulating nucleic acids (CNAs) and cancer – A survey. *Biochim Biophys Acta, Rev Cancer* 2007;1775:181–232.
- [6] Swarup V, Rajeswari MR. Circulating (cell-free) nucleic acids – a promising, non-invasive tool for early detection of several human diseases. *FEBS Lett* 2007;581:795–9.
- [7] Diehl F, Li M, Dressman D, et al. Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc Natl Acad Sci U S A* 2005;102:16368–73.
- [8] Juergens RA, Wrangle J, Vendetti FP, et al. Combination epigenetic therapy has efficacy in patients with refractory advanced non-small cell lung cancer. *Cancer Discov* 2011;1:598–607.
- [9] Baylin SB, Jones PA. A decade of exploring the cancer epigenome – biological and translational implications. *Nat Rev Cancer* 2011;11:726–34.
- [10] Levenson VV, Melnikov AA. DNA methylation as clinically useful biomarkers—light at the end of the tunnel. *Pharmaceuticals* 2012;5:94–113.
- [11] Herman JG. Circulating methylated DNA. In: Hoon DSB, Taback B, editors. *Circulating nucleic acids in plasma/serum iii and serum proteomics*. New York: New York Acad Sciences; 2004. p. 33–9.
- [12] Zeerleder S. The struggle to detect circulating DNA. *Crit Care* 2006;10.
- [13] Esteller M, Sanchez-Cespedes M, Rosell R, Sidransky D, Baylin SB, Herman JG. Detection of aberrant promoter hypermethylation of tumor suppressor genes in serum DNA from non-small cell lung cancer patients. *Cancer Res* 1999;59:67–70.
- [14] Schrader C, Schielke A, Ellerbroek L, John R. PCR inhibitors – occurrence, properties and removal. *J Appl Microbiol* 2012;113:1014–26.
- [15] Bailey VJ, Zhang Y, Keeley BP, et al. Single-tube analysis of DNA methylation with silica superparamagnetic beads. *Clin Chem* 2010;56:1022–5.
- [16] Cairns P. Gene methylation and early detection of genitourinary cancer: the road ahead. *Nat Rev Cancer* 2007;7:531–43.
- [17] Brock MV, Hooker CM, Ota-Machida E, et al. DNA methylation markers and early recurrence in stage I lung cancer. *N Engl J Med* 2008;358:1118–28.

- [18] Hesson LB, Cooper WN, Latif F. The role of RASSF1A methylation in cancer. *Dis Markers* 2007;23:73–87.
- [19] Wang J, Wang BC, Chen X, Bi JW. The prognostic value of RASSF1A promoter hypermethylation in non-small cell lung carcinoma: a systematic review and meta-analysis. *Carcinogenesis* 2011;32:411–6.
- [20] Daimon K, Komai S, Takarada Y. Method of extracting nucleic acids using particulate carrier. Japan: Toyo Boseki Kabushiki Kaisha; 2003.
- [21] Otto P. MagneSil™ paramagnetic particles: magnetics for DNA purification. *J Assoc Lab Autom* 2002;7:34–7.
- [22] Tan WH, Wang KM, He XX, et al. Bionanotechnology based on silica nanoparticles. *Med Res Rev* 2004;24:621–38.
- [23] Gallagher ML, Burke WF, Orzech K. Carrier RNA enhancement of recovery of DNA from dilute-solutions. *Biochem Biophys Res Commun* 1987;144:271–6.
- [24] Shi RB, Wang YC, Hu YL, Chen L, Wan QH. Preparation of magnetite-loaded silica microspheres for solid-phase extraction of genomic DNA from soy-based foodstuffs. *J Chromatogr A* 2009;1216:6382–6.
- [25] Ramirez JL, Sarries C, de Castro PL, et al. Methylation patterns and K-ras mutations in tumor and paired serum of resected non-small-cell lung cancer patients. *Cancer Lett* 2003;193:207–16.
- [26] Usadel H, Brabender J, Danenberg KD, et al. Quantitative adenomatous polyposis coli promoter methylation analysis in tumor tissue, serum, and plasma DNA of patients with lung cancer. *Cancer Res* 2002;62:371–5.

# Clinical Cancer Research



## Functional Identification of Cancer-Specific Methylation of *CDO1*, *HOXA9*, and *TAC1* for the Diagnosis of Lung Cancer

John Wrangle, Emi Ota Machida, Ludmila Danilova, et al.

*Clin Cancer Res* 2014;20:1856-1864. Published OnlineFirst January 31, 2014.

**Updated version** Access the most recent version of this article at:  
doi:[10.1158/1078-0432.CCR-13-2109](https://doi.org/10.1158/1078-0432.CCR-13-2109)

**Supplementary Material** Access the most recent supplemental material at:  
<http://clincancerres.aacrjournals.org/content/suppl/2014/02/04/1078-0432.CCR-13-2109.DC1.html>

**Cited Articles** This article cites by 35 articles, 12 of which you can access for free at:  
<http://clincancerres.aacrjournals.org/content/20/7/1856.full.html#ref-list-1>

**E-mail alerts** [Sign up to receive free email-alerts](#) related to this article or journal.

**Reprints and Subscriptions** To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at [pubs@aacr.org](mailto:pubs@aacr.org).

**Permissions** To request permission to re-use all or part of this article, contact the AACR Publications Department at [permissions@aacr.org](mailto:permissions@aacr.org).



## Functional Identification of Cancer-Specific Methylation of *CDO1*, *HOXA9*, and *TAC1* for the Diagnosis of Lung Cancer

John Wrangle<sup>1</sup>, Emi Ota Machida<sup>1</sup>, Ludmila Danilova<sup>1</sup>, Alicia Hulbert<sup>1</sup>, Noreli Franco<sup>1</sup>, Wei Zhang<sup>1</sup>, Sabine C. Glöckner<sup>1</sup>, Mathewos Tessema<sup>5</sup>, Leander Van Neste<sup>2,3</sup>, Hariharan Easwaran<sup>1</sup>, Kornel E. Schuebel<sup>1</sup>, Julien Licchesi<sup>1</sup>, Craig M. Hooker<sup>1</sup>, Nita Ahuja<sup>1</sup>, Jun Amano<sup>4</sup>, Steven A. Belinsky<sup>5</sup>, Stephen B. Baylin<sup>1</sup>, James G. Herman<sup>1</sup>, and Malcolm V. Brock<sup>1</sup>

### Abstract

**Purpose:** Non-small cell lung cancer (NSCLC) is the leading cause of cancer mortality in the world. Novel diagnostic biomarkers may augment both existing NSCLC screening methods as well as molecular diagnostic tests of surgical specimens to more accurately stratify and stage candidates for adjuvant chemotherapy. Hypermethylation of CpG islands is a common and important alteration in the transition from normal tissue to cancer.

**Experimental Design:** Following previously validated methods for the discovery of cancer-specific hypermethylation changes, we treated eight NSCLC cell lines with the hypomethylating agent deoxyazacitidine or trichostatin A. We validated the findings using a large publicly available database and two independent cohorts of primary samples.

**Results:** We identified >300 candidate genes. Using The Cancer Genome Atlas (TCGA) and extensive filtering to refine our candidate genes for the greatest ability to distinguish tumor from normal, we define a three-gene panel, *CDO1*, *HOXA9*, and *TAC1*, which we subsequently validate in two independent cohorts of primary NSCLC samples. This three-gene panel is 100% specific, showing no methylation in 75 TCGA normal and seven primary normal samples and is 83% to 99% sensitive for NSCLC depending on the cohort.

**Conclusion:** This degree of sensitivity and specificity may be of high value to diagnose the earliest stages of NSCLC. Addition of this three-gene panel to other previously validated methylation biomarkers holds great promise in both early diagnosis and molecular staging of NSCLC. *Clin Cancer Res*; 20(7); 1856–64. ©2014 AACR.

### Introduction

Non-small cell lung cancer (NSCLC) is the leading cause of cancer-related mortality worldwide (1, 2). Although improvements in the treatment of advanced stage lung malignancies have been made, including agents targeting specific genetic aberrations, epigenetic therapies, and

exploiting the potential of the immune system to assert control over tumor growth, lung cancer remains the main cause of cancer-related deaths (3–5). Cancer-specific molecular changes have utility not only as targets for therapy, but also as biomarkers for the determination of risk of recurrence for early-stage lung cancer. Such prognostic capability may be due to the biologic significance of the alteration or because detection of molecular alterations in lymph nodes may herald a higher stage of disease than is detectable by current pathology standards (6, 7).

There is also much interest in early detection strategies to improve outcomes in lung cancer, which have culminated in the landmark National Lung Screening Trial (NLST). Although the 20% relative reduction in lung cancer mortality in the NLST low-dose computed tomography (CT) screening arm is encouraging, it belies a false positive rate among screening results of 96.4%, which has resulted in some pause among clinicians and payers alike for immediate widespread adoption of the technique (8). Improved techniques or ancillary testing methods to augment both the sensitivity and specificity of screening for lung cancer could augment CT screening.

**Authors' Affiliations:** <sup>1</sup>The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, Maryland; <sup>2</sup>Department of Molecular Biotechnology, Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium; <sup>3</sup>MDxHealth Inc, Irvine, California; <sup>4</sup>Shinshu University School of Medicine, Asahi, Matsumoto, Nagano, Japan; and <sup>5</sup>Lovelace Respiratory Research Institute, Albuquerque, New Mexico

**Note:** Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

J. Wrangle and E.O. Machida contributed equally to this work.

**Corresponding Author:** Malcolm V. Brock, Cancer Biology Program, The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, 1650 Orleans St., Baltimore, MD 21231. Phone: 410-955-8506; Fax: 410-614-9884; E-mail: mbrock1@jhmi.edu

doi: 10.1158/1078-0432.CCR-13-2109

©2014 American Association for Cancer Research.

### Translational Relevance

Lung cancer remains the leading cause of cancer-related mortality in the world. The likelihood of mortality related to the disease increases dramatically with the stage of disease. Using a validated experimental method of eliciting frequently methylated genes in cancer, which we then examine in hundreds of lung cancer samples in The Cancer Genome Atlas and two, independent cohorts, we describe DNA methylation of one or more of *CDO1*, *HOXA9*, and *TAC1* as nearly universal in lung cancer in the United States. Such a highly sensitive and specific molecular marker of disease may play a significant role in improving early detection strategies and decreasing NSCLC morbidity and mortality.

The most promising nonradiologic ancillary tests involve the detection of cancer-specific events in tissues or fluids carrying tumor cells or tumor DNA, such as lymph node samples, sputum, or plasma. Because cancer-specific DNA methylation events are common and occur early in lung cancer progression, recent studies have used nested methylation-specific PCR (MSP) for detection of promoter methylation in sputum (9, 10). For example, using *PAX5α*, *GATA5*, and *SULF2* genes derived from studies of genes with known biologic importance in NSCLC demonstrated the ability to predict the outcome of a diagnosis of lung cancer in two high-risk cohorts (11–14). Although these studies demonstrate the feasibility of molecular detection of altered, cancer-specific DNA methylation in sputum, there remains a need for improvement in the panel of markers used. The measure of success expected from a test lies in the frequency of the event (sensitivity) and the absence of the event in normal samples (specificity). In this work, we seek to build upon approaches that define the most highly sensitive and specific markers of cancer, which have often been found to be linked to polycomb-associated sites in embryonic stem cells, toward the deployment of a clinically useful assay (15–17). We hypothesized that the current genes used in strategies to assess presence or absence of lung cancer based on sputum and other bodily tissues and fluids may be augmented by a method combining preclinical and population-based studies to identify the most highly sensitive and specific methylation events in lung cancer.

Here, we report the discovery and characterization of genomic changes in DNA methylation occurring in association with a described biologic program, moving from the study of individual loci to a comprehensive analysis of alterations in NSCLC with the intention of uncovering epigenetic events which may predict a cancer's natural history or be utilized for the molecular detection of disease. This study provides a method for systematic discovery of epigenetic biomarkers which may be used for improving the screening and diagnosis of this deadly disease.

### Materials and Methods

#### Cell culture and treatment

All NSCLC cell lines were purchased from the American Type Culture Collection. H838, H23, H1993, H1568, H2170, and H520 were cultured in RPMI-1640 medium (Mediatech, Inc.); H1869 was cultured in DMEM/F-12 Medium and SK-MES-1 was cultured in Dulbecco's Modified Eagle Medium (DMEM; Mediatech, Inc.). Cell lines H838, H23, H1993, and H1568 were derived from adenocarcinomas and H2170, H520, H1869, and SK-MES-1 were derived from squamous cell carcinomas. Cell lines of squamous carcinoma and adenocarcinoma histology are represented equally so that cancer-specific, rather than histology-specific markers may be elicited by the experimental method. All cell culture media were supplemented with 10% bovine calf serum (BCS) and incubated in humidified air and 5% CO<sub>2</sub> at 37°C. For drug treatments, log phase cells were cultured in growth media containing 10% BCS and 1 × penicillin/streptomycin with 5 μmol/L decitabine (Sigma; stock solution: 1 mmol/L in PBS) for 96 hours, replacing fresh media and decitabine every 24 hours. Cell treatment with 300 nmol/L Trichostatin A (TSA; Sigma; stock solution: 1.5 mmol/L dissolved in ethanol) was performed for 18 hours. Control cells underwent mock treatment in parallel with addition of equal volumes of PBS or ethanol without drugs.

#### Microarray analysis

RNA was harvested from cells in log phase growth using TRIzol (Invitrogen) and the RNeasy kit with DNase digestion (Qiagen) according to the manufacturer's instructions. RNA was quantified using the NanoDrop ND-100 followed by quality assessment with the 2100 Bioanalyzer (Agilent Technologies). RNA concentrations for each sample was more than 200 ng/μL, with 28S/18S ratios more than 2.2 and RNA integrity scores of 10 (10 scored as the highest). Sample amplification and labeling procedures were carried out using the Low RNA Input Fluorescent Linear Amplification Kit (Agilent Technologies). The labeled cRNA was purified using the RNeasy Mini Kit (Qiagen) and quantified. RNA spike-in controls (Agilent Technologies) were added to RNA samples before amplification. Samples (0.75 μg) labeled with Cy3 or Cy5 were mixed with control targets (Agilent Technologies), assembled on Oligo Microarray, hybridized, and processed according to the Agilent microarray protocol. Scanning was performed with the Agilent G2505B microarray scanner using settings recommended by Agilent Technologies. Microarray data are available in the ArrayExpress database under accession number E-MTAB-1939.

#### Data analysis for microarray

Quality checks for all arrays included visual inspection for artifacts and the distribution of signal and background intensity for red and green channels. All arrays passed quality checks and were used. The statistical platform R and packages from Bioconductor were used for all computation (18, 19). The log ratio of red signal to green signal was calculated after background subtraction and

LoEss normalization as implemented in the limma package from Bioconductor (20). Individual arrays were scaled to have the same interquartile range (75th percentile–25th percentile).

### Methylation and gene expression analysis

RNA was isolated with TRIzol Reagent (Invitrogen) according to the manufacturer's instructions. For real-time PCR (RT-PCR), 1 µg of total RNA was reverse transcribed using SuperScript First-Strand Synthesis System for RT-PCR (Invitrogen). For MSP analysis, DNA was extracted following a standard phenol-chloroform extraction method. Bisulfite modification of genomic DNA was carried out using the EZ DNA methylation Kit (Zymo Research). Primer sequences specific to unmethylated and methylated promoter sequences were designed using MSPPrimer (21). MSP was performed as previously described (22). Ten microliters of all PCR products were loaded directly onto 2% agarose gels containing GelStar Nucleic Acid Gel Stain (Cambrex Corp.) and visualized under UV illumination. Primer sequences and conditions for MSP are available upon request.

### Human tissue analysis

Fifty-nine primary lung cancers were obtained from Johns Hopkins Hospital in Baltimore, MD (Cohort A) and 30 from Shinshu University Hospital in Matsumoto, Japan (Cohort B). All tissues were immediately frozen at  $-80^{\circ}\text{C}$  after surgical resection. Normal lung cDNA was purchased from DNA Technologies Inc. Six normal lung tissues were obtained from individuals without cancer (five from autopsy and one from lung peripheral to a benign bronchial tumor). Tissue acquisition was conducted under approved guidelines of the Institutional Review Boards from both institutions. Histologic examination was based on World Health Organization classification criteria (23). Clinical staging was done according to Mountain and Dresler's tumor-node-metastasis classification criteria (24).

### TCGA analysis data and methods

We used the DNA methylation data of 409 lung adenocarcinoma samples with 32 matched normal samples as well as 227 lung squamous cell carcinoma samples with 43 matched normal samples from the Cancer Genome Atlas (TCGA) project (25, 26). DNA methylation was measured on the Illumina HumanMethylation 450 K platform (18, 27).

The analysis of DNA methylation data was performed using R/Bioconductor software with the limma package and custom routines for data analysis (18, 19, 28). We selected only those probes for sites situated within CpG-island promoters of genes unmethylated at their promoter sites in all normal TCGA samples ( $\beta$ -value  $< 0.2$ ). For each probe we estimated a  $t$  statistic and  $P$  value by fitting a linear model of its differential methylation between tumor and normal samples (29). All probes tested had adjusted  $P$  values less than  $1 \times 10^{-4}$ . Figure 1 shows a heatmap of DNA methylation level for each site (in rows) for all tumor and normal

samples (in columns). The columns of the heatmap were ordered by unsupervised clustering, whereas rows were ordered top-to-bottom by decreasing value of significance for  $t$  statistic for differential methylation. The sites and corresponding statistics for all probes can be found in Supplementary Table S1.

### Clustering analysis

DNA methylation clusters were based on the most variable CpG sites from Fig. 1 and on stage I and II samples. Consensus clustering was applied as implemented in the Bioconductor package ConsensusClusterPlus, with Euclidean distance and partitioning around medoids (pam) was used to derive clusters (30, 31).

### Survival analyses

$P$  value was computed from the Cox regression (the *coxph* function of the survival package; refs. 32, 33). Kaplan–Meier curves were made with the help of the *survfit* function from the same package using TCGA data for stage I and II tumors. The clinical endpoint for analysis was time to death. TCGA samples are not annotated for therapies received; therefore, no control for treatment in analysis is possible but may be assumed to represent the standard of care in the United States. Methylation data were obtained by TCGA from fresh-frozen tumors examined by Infinium HumanMethylation450 as previously described (25). Categorization for groups of comparison for survival outcomes is based on medoid clustering as described in Clustering Analysis.

### Binary DNA methylation assessment

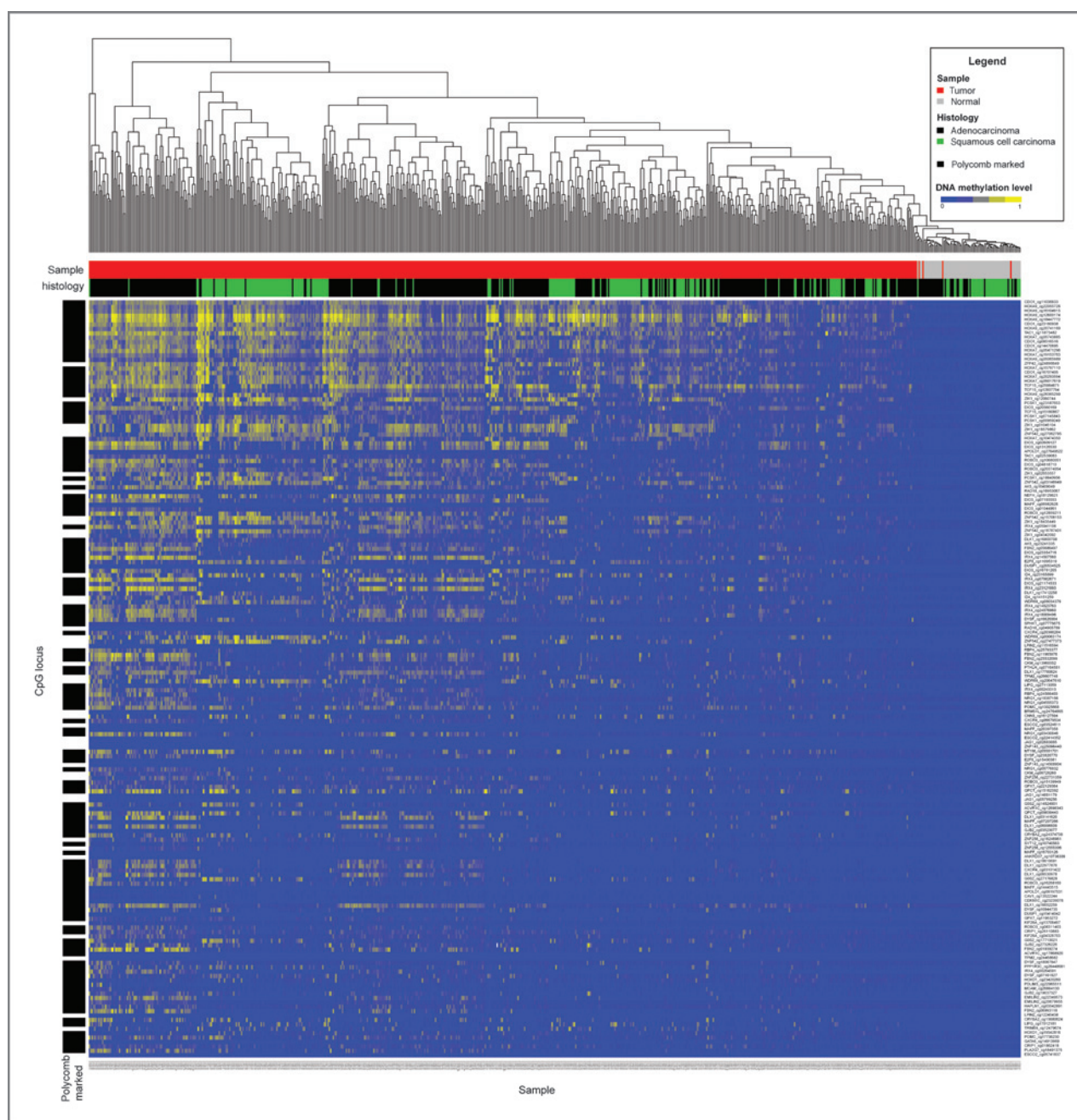
We selected the most significant CpG site per gene to define binary DNA methylation. For each gene, a sample was labeled DNA hypermethylated if the individual  $\beta$ -value of the gene was more than three times the SD of the mean of all combined  $\beta$ -values of normal samples.

## Results

### Functional identification of cancer-specific, hypermethylated genes in NSCLC cell lines

On the basis of a previously designed method to unmask epigenetically silenced cancer-specific, DNA-hypermethylated genes, we treated eight NSCLC cell lines with either the DNA-methylation and DNMT inhibitor, decitabine, or the histone deacetylase (HDAC) class I/II histone deacetylase inhibitor, TSA (34, 35). Gene expression changes determined using Affymetrix microarray for decitabine- or TSA-treated cells were compared with mock-treated cells. This method enables the identification of genes induced specifically by decitabine, an important distinction as decitabine has the capacity to induce gene reexpression of loci silenced predominantly by hypermethylation, whereas TSA alone will fail to induce reexpression (34). The objective of methylation biomarker discovery by decitabine-specific reexpression is to generate a list of genes likely to be silenced by methylation of promoter CpG islands. Decitabine-specific reexpression for a gene is defined as a more than 2.0-fold reexpression on a microarray with decitabine treatment





**Figure 1.** Cancer-specific DNA methylation discriminates NSCLC tumors from normal lung samples. Methylation data are derived from 636 NSCLC in TCGA representing 227 lung squamous carcinomas with 43 matched normal samples and 409 lung nonsquamous carcinomas with 32 matched normal samples. Columns represent tumor or normal tissue samples. Rows represent individual methylation probes from the Infinium methylation array. The ability of each probe to discriminate tumor versus normal and an associated  $t$  statistic was estimated by a linear model for each CpG island promoter probe. Only probes with significant  $P$  values are included in the heatmap. Rows are ordered from top-to-bottom by  $P$  value. All  $P$  values are  $<0.0001$ . Probes with mean  $\beta$ -values  $>0.2$  in normal samples were excluded from the analysis. Of the 305 genes exhibiting decitabine-specific upregulation, 63 genes represented by 172 methylation probes met the preceding criteria. Columns are ordered by unsupervised hierarchical clustering. A few tumors cluster with normal samples. This is consistent with prior TCGA analyses that demonstrate "normal-like" methylation patterns in a subset of tumors.

compared with mock-treated cells, less than 1.4-fold reexpression with TSA treatment compared with mock-treated cells, and no basal expression in mock-treated cells as previously described (34, 35). To find genes which would be expected to have higher frequencies of methylation in

lung cancer, we refined this list to require the preceding criteria in at least two of eight cell lines. A total of 305 genes were determined to be upregulated by decitabine using these criteria from eight NSCLC cell lines (Supplementary Fig. S1).

### Refining a diagnostic three-gene panel of cancer-specific, hypermethylated genes in NSCLC using The Cancer Genome Atlas dataset

The comprehensive analysis of 305 genes in primary tumors to determine their utility would represent a challenging task without additional informatics filters to select the most promising candidates. To refine this list of genes, we applied this functionally derived gene list to primary tumors characterized in the TCGA lung cancer project, and then validated the findings in two, independent single-institution cohorts of primary NSCLC tumors (Table 1). We first tested for tumor specificity among the TCGA tumors, comparing DNA methylation between lung tumors and normal lung tissue. Of the 305 decitabine upregulated genes, 63 genes with a total of 172 annotated CpG island promoter probes on the Infinium 450 K array had a statistically significant ability to differentiate tumor versus normal in TCGA samples as estimated by a linear regression model. In addition, these genes had extremely low methylation ( $\beta$ -values) in TCGA normal samples, thereby defining a group of decitabine-responsive, cancer-specific methylated genes. Data using these probes are represented in a heatmap where rows are ordered from top to bottom by *P* values based on the ability of an individual methylation array probe to distinguish tumor versus normal. Columns are ordered by unsupervised hierarchical clustering (Fig. 1 and Supplementary Table S1). Maximum estimated *P* value for each probe was  $1 \times 10^{-4}$ . *CDO1*, *HOXA9*, and *TAC1* were notable for extremely high rates of DNA methylation

in tumors and low methylation in normal samples, and were most effective in distinguishing tumor versus normal based on *P* value of linear logistic regression model.

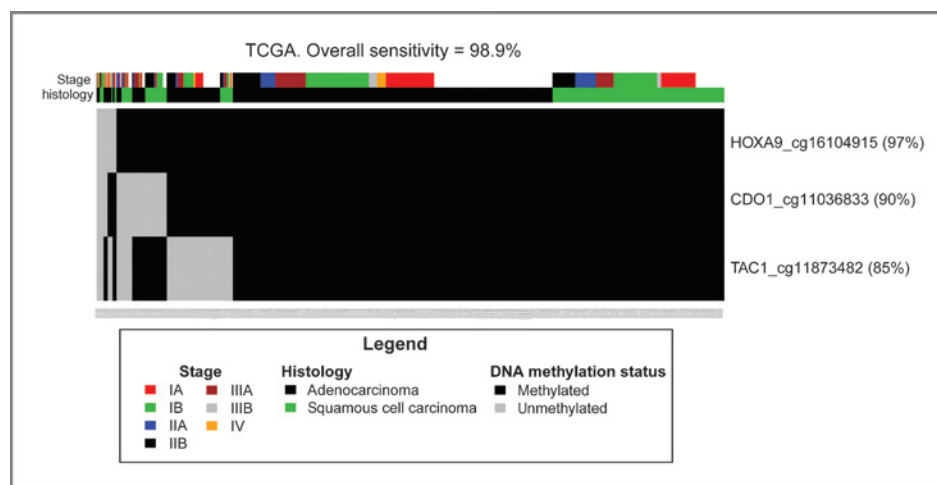
Binary methylation values as determined by the single best methylation probe from the promoter CpG islands of *CDO1*, *HOXA9*, and *TAC1*, and were plotted for all NSCLC stages together as well as for stage I alone (Fig. 2 and Supplementary Fig. S2 and Supplementary Table S1). Sensitivity is not limited by histology or tumor stage in the TCGA dataset. In fact, methylation of at least one of these three genes is 98.9% sensitive for tumors stage I–IV and 98.7% sensitive for stage I tumors alone. *HOXA9* alone is methylated in 97% of NSCLC TCGA samples. There are limited descriptions of DNA methylation of these genes in human lung cancer in previous studies. Although *TAC1* promoter methylation has not been described in lung malignancies, highly prevalent *HOX* cluster gene methylation, including *HOXA9*, has been reported in cell lines and a small number of squamous stage I tumors ( $n = 4$ ) as well as a pool of mixed stage and mixed histology tumors ( $n = 20$ ; refs. 17, 36). *HOXA9* hypermethylation has been described as a potential screening test in combination with *SOX1* hypermethylation and *DDR1* hypomethylation as assayed by pyrosequencing (37). *CDO1* has been reported as a methylated gene in squamous lung tumors ( $n = 30$ ; ref. 38). *CDO1* and *TAC1* have been described as high-prevalence cancer-specific methylated genes in breast cancer (35). However, no previous study has described the sensitivity and

**Table 1.** Clinicopathological characteristics of patient cohorts

		Cohort		
		TCGA	A	B
		(n = 636)	(n = 59)	(n = 30)
Age	Average, y	68	65.8	64.1
Sex	F (%)	238 (37.4%)	27 (45.8%)	11 (36.7%)
	M (%)	306 (48.1%)	32 (54.2%)	19 (63.3%)
	NA	92 (14.5%)	0	0
Smoking	Ever	466 (73.3%)	47 (79.7%)	NA
	Never	61 (9.6%)	4 (6.8%)	NA
	NA	109 (17.1%)	8 (13.6%)	NA
Histology	Adeno	409 (64.3%)	36 (61.0%)	21 (70%)
	SCC	227 (35.7%)	23 (39.0%)	9 (30%)
Stage	Ia	125 (19.7%)	16 (27.1%)	3 (10%)
	Ib	159 (25.0%)	20 (33.9%)	4 (13.3%)
	IIa	58 (9.1%)	1 (1.7%)	3 (10%)
	IIb	84 (13.2%)	9 (15.3%)	6 (20%)
	IIIa	78 (12.2%)	7 (11.9%)	7 (23.3%)
	IIIb	14 (2.2%)	3 (5.1%)	4 (13.3%)
	IV	17 (2.7%)	3 (5.1%)	3 (10%)
	NA	101 (15.9%)	0	0

NOTE: TCGA is a publicly available database that contains DNA methylation data for hundreds of primary patients with NSCLC. Cohort A consists of resected patients with NSCLC from Johns Hopkins Hospital in Baltimore, MD. Cohort B consists of resected patients with NSCLC from Shinshu University Hospital in Matsumoto, Japan.

**Figure 2.** DNA methylation of *CDO1*, *HOXA9*, and *TAC1* is highly sensitive for NSCLC in TCGA. A single Infinium methylation probe with the best discriminative capacity between tumor and normal samples was selected for each of the three genes. A sample is considered methylated for a gene if its  $\beta$ -value was larger than three times the SD of the mean of  $\beta$ -values of normal samples. Methylation of at least one gene-promoter among *CDO1*, *HOXA9*, and *TAC1* by Infinium array identifies 98.9% of NSCLC cases in 636 cases in TCGA.



specificity for a combination of these genes in a large population of NSCLC tumors and validation cohorts.

In addition to their diagnostic utility, we examined the potential prognostic significance of this functionally derived list of cancer-specific methylation. As would be expected from a list of genes with an extremely high prevalence of methylation and no described biologic role in lung cancer, none of the 63 genes examined individually was associated with survival outcome in TCGA (data not shown). To examine whether methylation of these genes taken as a group reflects biologic differences in tumors, we clustered all TCGA lung cancer samples using medoid clustering, a method for defining optimal numbers of groups within a dataset. When taken together, the 63 cancer-specific hypermethylated genes form three groups, adenocarcinoma-predominant, squamous-predominant, and a mixed group. These clusters demonstrate a marginal association with survival in the TCGA tumors ( $P = 0.04$ ; Supplementary Fig. S3). From our previously published markers of outcome in early-stage, resected lung cancer, our strongest associations with outcome came from questions pertaining to cancer-specific methylation confirmed in lymph nodes, thus a diagnostic or staging paradigm. As the TCGA contains only samples of primary tumors and no associated lymph nodes, there is no ability to assess concordance of methylation between tumor and lymph node. When examining tumor-only questions from our previous work, we find general agreement with the moderate prognostic capacity of methylation of four genes when examined in tumor only, highlighting the need to refine a highly sensitive and specific diagnostic markers for the molecular staging of NSCLC (ref. 6; Supplementary Fig. S4).

#### Association of progenitor cell polycomb-associated genes with cancer-specific methylation marks

Previous studies have suggested that genes with polycomb marks in chromatin surrounding the transcription start sites are predisposed to aberrant DNA methylation silencing in cancer (15, 39, 40). In embryonic stem cells, polycomb association occurs in the context of bivalent

chromatin marks containing both active histone 3 lysine 4 trimethylation (H3K4me3) and repressive histone 3 lysine 27 trimethylation (H3K27me3) marks. Of the 63 cancer-specific hypermethylated genes, 45 (71.4%) are considered bivalent genes silenced by polycomb-repressive complex in progenitor cell states, a rate much higher than the presence of these marks among all genes (21% using estimated 4,413 bivalent genes among an estimated 21,000 total human genes,  $P < 0.0001$ ; refs. 15, 38). *CDO1*, *HOXA9*, and *TAC1* are all polycomb associated in embryonic stem cells (Fig. 1 and Supplementary Table S1).

#### Validating the diagnostic utility of a three-gene panel in two cohorts of primary tissue

To confirm the high prevalence of DNA methylation for these genes in other primary lung tumors, we then validated the sensitivity of these three genes in two independent cohorts of NSCLC tumor samples using MSP (Table 1; Fig. 3). Primers for *CDO1*, *HOXA9*, and *TAC1* were designed and tested on tumor samples from cohorts in the United States and Japan. As was observed for these genes on the Infinium platform within TCGA data, there was no methylation in seven normal lung samples when examined using MSP. In contrast with normal lung, among the American cohort A and Japanese cohort B, respectively, 94.9% and 83.3% of the tumor samples were methylated for at least one of these three genes. Because this three-gene panel has near-zero methylation  $\beta$ -values by Infinium and MSP in normal tissues and is found to have stage-independent hypermethylation in cancer, these genes fulfill critical characteristics for designing a threshold for methylation in clinical assays and for identifying the earliest stages of NSCLC (Fig. 3).

#### Discussion

Using an experimental model to derive a list of candidate cancer-specific, hypermethylated, polycomb-associated genes in lung cancer, we validated a three-gene test in a large publicly available database and two independent cohorts to describe a highly sensitive, highly specific



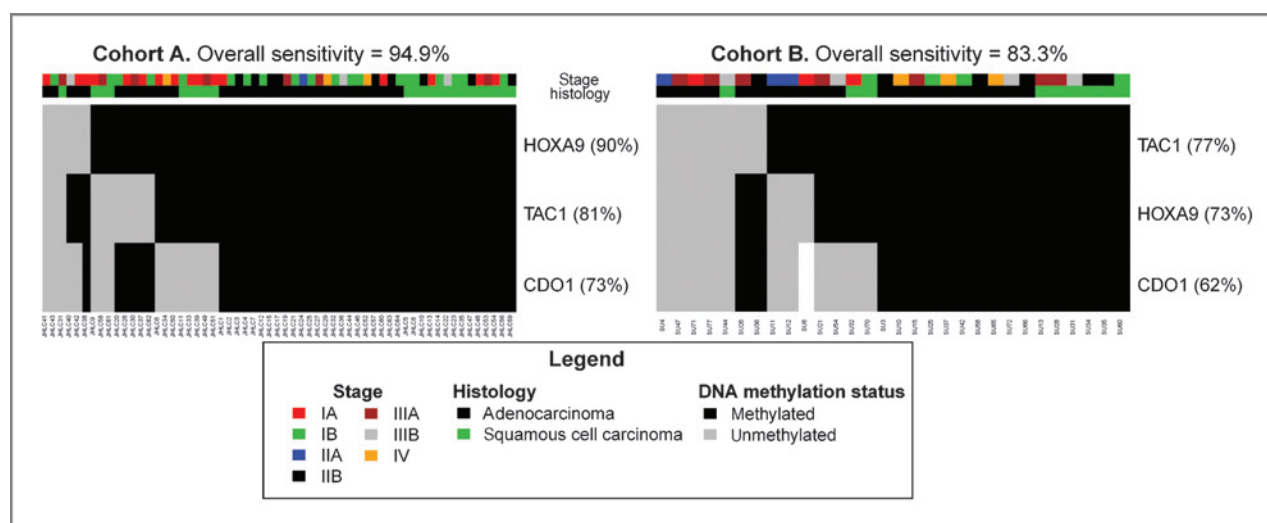


Figure 3. Validation of the sensitivity of methylation-specific PCR for three genes in two independent cohorts. Methylation of at least one gene-promoter among *CDO1*, *HOXA9*, or *TAC1* by MSP identifies 94.9% of NSCLC cases in 59-patient United States cohort A and 83.3% of NSCLC cases from the independent 30-patient Japanese cohort B.

diagnostic test for NSCLC. In the present study, we use a functional approach to identify three genes, *CDO1*, *HOXA9*, and, *TAC1*, in which we describe cancer-specific DNA methylation without regard for the biologic implication of that cancer-specific methylation. When examining diagnostic sensitivity, we find a remarkable concordance between TCGA samples, derived entirely from American hospitals, and our American validation cohort with sensitivities of 98.9% and 94.9%, respectively. Diagnostic sensitivity in the Japanese cohort is similar but lower at 83.3%. Although some variation may be due to sampling, we can also reasonably hypothesize that this reflects other established differences in the NSCLC populations of American and Japan and highlights the need to tailor a test precisely to target populations. Although an 83% sensitivity of detection far exceeds any mutational detection approach currently available, it may be possible to provide an even better three-gene test if these genes were chosen from among highly methylated genes determined from analysis of lung cancers in Japanese populations.

In addition, we have explored whether these cancer-specific alterations may have prognostic value. As might be expected, these genes without an established role in the pathogenesis of lung cancer and/or an extremely high prevalence of methylation prove to be of no prognostic value when examined individually. Indeed, in our previously published study of four genes, there was limited prognostic value when knowledge of methylation status is known for the tumor only. In addition, our previous study suggested that the presence of cancer-specific methylation in histologically negative lymph nodes, particularly mediastinal (N2) nodes, was most prognostic of recurrence and lung cancer associated (6).

An interesting characteristic of the genes elicited by this functional screen for novel cancer-specific biomarkers is a high degree of overlap with polycomb-associated genes.

H3K4me3 and H3K27me3 define a bivalent chromatin state that denotes a low-transcriptional, poised state for a group of genes in progenitor and stem cells highly enriched for developmental processes (41). These genes, largely active during development of differentiated tissues, are downregulated by the polycomb-repressive complex when a chromatin bivalent state exists and are largely devoid of DNA methylation. These loci are particularly vulnerable to DNA methylation during the process of carcinogenesis (15). Although the mechanism that underlies epigenetic silencing transitioning from the polycomb-repressive complex to DNA methylation would suggest little or no alteration in gene expression in some cases, assaying these methylation changes remains useful as highly sensitive and specific hallmarks of tumor tissue and are therefore excellent candidates as diagnostic biomarkers. In addition, because different stem and progenitor populations show variation in distribution of chromatin-bivalency, the methylation marks at polycomb-associated DNA may signal subtle differences in the cell of origin.

For the molecular detection of disease in lymph nodes for staging and for approaches for early detection involving sputum, plasma, or fine needle aspirates, molecular alterations present in the vast majority of tumors will be the most sensitive and efficient means of detection. Through the characterization of hypermethylated loci reported here, we have developed a highly sensitive, highly specific test for identifying cases of NSCLC which may serve these purposes. A three-gene methylation assay with sensitivity in tumors approaching 100% may allow for the detection or diagnosis of disease in tissues remote from the primary tumor without specific knowledge of those genes in the tumor itself. The present study demonstrates the performance of a three-gene test in primary tumor samples for which inadequate diagnostic methods currently exist. With improvements in detection of DNA methylation in blood

and sputum, the sensitivity of detection in additional types of biospecimens, including plasma and sputum samples, can now be tested (42).

### Disclosure of Potential Conflicts of Interest

L.V. Neste is an inventor and has ownership interest in a patent for the detection and prognosis of lung cancer. S.B. Baylin has received speakers bureau honoraria from MDxHealth and is a consultant/advisory board member for Bionumberick Pharm. J.G. Herman has commercial research support from and is a consultant/advisory board member for MDxHealth. No potential conflicts of interest were disclosed by the other authors.

### Authors' Contributions

**Conception and design:** J. Wrangle, E.O. Machida, A. Hulbert, L.V. Neste, J.G. Herman

**Development of methodology:** J. Wrangle, E.O. Machida, A. Hulbert, N. Franco, L.V. Neste, K.E. Schuebel, N. Ahuja

**Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.):** J. Wrangle, E.O. Machida, A. Hulbert, N. Franco, W. Zhang, M. Tessema, J. Licchesi

**Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** J. Wrangle, E.O. Machida, L. Danilova, A. Hulbert, S. Glockner, L.V. Neste, C.M. Hooker, J.G. Herman

**Writing, review, and/or revision of the manuscript:** J. Wrangle, E.O. Machida, L. Danilova, M. Tessema, L.V. Neste, N. Ahuja, J. Amano, S.A. Belinsky, S.B. Baylin, J.G. Herman, M.V. Brock

**Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases):** E.O. Machida, N. Franco, L.V. Neste, H. Easwaran, N. Ahuja, J. Amano

**Study supervision:** N. Ahuja, J.G. Herman

### Acknowledgments

The authors thank NIH Lung SPORE CA058184, DOD W81XWH-12-1-0323, and MDxHealth for support and also Leslie Cope and Mary Lou Wrangle.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received August 5, 2013; revised November 19, 2013; accepted November 30, 2013; published OnlineFirst January 31, 2014.

### References

- Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. *CA Cancer J Clin* 2013;63:11–30.
- Youlten DR, Cramb SM, Baade PD. The International Epidemiology of Lung Cancer: geographical distribution and secular trends. *J Thorac Oncol* 2008;3:819–31.
- Vadakara J, Borghaei H. Personalized medicine and treatment approaches in non-small-cell lung carcinoma. *Pharmgenomics Pers Med* 2012;5:113–23.
- Juergens RA, Wrangle J, Vendetti FP, Murphy SC, Zhao M, Coleman B, et al. Combination epigenetic therapy has efficacy in patients with refractory advanced non-small cell lung cancer. *Cancer Discov* 2011;1:598–607.
- Reck M. What future opportunities may immuno-oncology provide for improving the treatment of patients with lung cancer? *Ann Oncol* 2012;23 Suppl 8:viii28–34.
- Brock MV, Hooker CM, Ota-Machida E, Han Y, Guo M, Ames S, et al. DNA methylation markers and early recurrence in stage I lung cancer. *N Engl J Med* 2008;358:1118–28.
- Beer DG, Kardia SL, Huang CC, Giordano TJ, Levin AM, Misek DE, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med* 2002;8:816–24.
- Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;365:395–409.
- Belinsky SA, Nikula KJ, Palmisano WA, Michels R, Saccomanno G, Gabrielson E, et al. Aberrant methylation of p16(INK4a) is an early event in lung cancer and a potential biomarker for early diagnosis. *Proc Natl Acad Sci U S A* 1998;95:11891–6.
- Palmisano WA, Divine KK, Saccomanno G, Gilliland FD, Baylin SB, Herman JG, et al. Predicting lung cancer by detecting aberrant promoter methylation in sputum. *Cancer Res* 2000;60:5954–8.
- Belinsky SA, Klinge DM, Dekker JD, Smith MW, Bocklage TJ, Gilliland FD, et al. Gene promoter methylation in plasma and sputum increases with lung cancer risk. *Clin Cancer Res* 2005;11:6505–11.
- Belinsky SA, Liechty KC, Gentry FD, Wolf HJ, Rogers J, Vu K, et al. Promoter hypermethylation of multiple genes in sputum precedes lung cancer incidence in a high-risk cohort. *Cancer Res* 2006;66:3338–44.
- Machida EO, Brock MV, Hooker CM, Nakayama J, Ishida A, Amano J, et al. Hypermethylation of ASC/TMS1 is a sputum marker for late-stage lung cancer. *Cancer Res* 2006;66:6210–8.
- Leng S, Do K, Yingling CM, Picchi MA, Wolf HJ, Kennedy TC, et al. Defining a gene promoter methylation signature in sputum for lung cancer risk assessment. *Clin Cancer Res* 2012;18:3387–95.
- Easwaran H, Johnstone SE, Van Neste L, Ohm J, Mosbrugger T, Wang Q, et al. A DNA hypermethylation module for the stem/progenitor cell signature of cancer. *Genome Res* 2012;22:837–49.
- Kim J, Woo AJ, Chu J, Snow JW, Fujiwara Y, Kim CG, et al. A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. *Cell* 143:313–24.
- Rauch T, Wang Z, Zhang X, Zhong X, Wu X, Lau SK, et al. Homeobox gene methylation in lung cancer studied by genome-wide analysis with a microarray-based methylated CpG island recovery assay. *Proc Natl Acad Sci U S A* 2007;104:5527–32.
- team RDC. R: A language and environment for statistical computing. R foundation for statistical computing. Vienna, Austria; 2004.
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004;5:R80.
- Smyth GK, Speed T. Normalization of cDNA microarray data. *Methods* 2003;31:265–73.
- Brandes JC, Carraway H, Herman JG. Optimal primer design using the novel primer design program: MSPprimer provides accurate methylation analysis of the ATM promoter. *Oncogene* 2007;26:6229–37.
- Herman JG, Graff JR, Myohanen S, Nelkin BD, Baylin SB. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci U S A* 1996;93:9821–6.
- Gibbs AR, Thunnissen FB. Histological typing of lung and pleural tumours: third edition. *J Clin Pathol* 2001;54:498–9.
- Mountain CF, Dresler CM. Regional lymph node classification for lung cancer staging. *Chest* 1997;111:1718–23.
- Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 2012;489:519–25.
- The Cancer Genome Atlas [database on the Internet]. Available from: <https://tcga-data.nci.nih.gov/>.
- Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. *Genomics* 2011;98:288–95.
- Smyth G. Limma: linear models for microarray data. In: Gentleman VC R., Dudoit S., Irizarry R., Huber W., editor. *Bioinformatics and Computational Biology Solutions using R and Bioconductor*. NY: Springer; 2005. p. 397–420.
- Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*; 2004;3:Article3. Epub 2004 Feb 12.
- Wilkerson M. ConsensusClusterPlus: ConsensusClusterPlus. R package version 1.5.1. 2011.



31. Reynolds A, Richards G, de la Iglesia B, Rayward-Smith V. Clustering rules: a comparison of partitioning and hierarchical clustering algorithms. *J Math Model Algorithms* 1992;5:475–504.
32. Andersen PaG R. Cox's regression model for counting processes, a large sample study. *Ann Stat* 1982;10:1033–1311.
33. Therneau T, Grambsch P. *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag; 2000.
34. Schuebel KE, Chen W, Cope L, Glockner SC, Suzuki H, Yi JM, et al. Comparing the DNA hypermethylome with gene mutations in human colorectal cancer. *PLoS Genet* 2007;3:1709–23.
35. Jeschke J, Van Neste L, Glockner SC, Dhir M, Calmon MF, Deregowski V, et al. Biomarkers for detection and prognosis of breast cancer identified by a functional hypermethylome screen. *Epigenetics* 2012;7:701–9.
36. Son JW, Jeong KJ, Jean WS, Park SY, Jheon S, Cho HM, et al. Genome-wide combination profiling of DNA copy number and methylation for deciphering biomarkers in non-small cell lung cancer patients. *Cancer Lett* 2011;311:29–37.
37. Nelson HH, Marsit CJ, Christensen BC, Houseman EA, Kontic M, Wiemels JL, et al. Key epigenetic changes associated with lung cancer development: results from dense methylation array profiling. *Epigenetics* 2012;7:559–66.
38. Kwon YJ, Lee SJ, Koh JS, Kim SH, Lee HW, Kang MC, et al. Genome-wide analysis of DNA methylation and the gene expression change in lung cancer. *J Thorac Oncol* 2012;7:20–33.
39. Noshmeh H, Weisenberger DJ, Diefes K, Phillips HS, Pujara K, Berman BP, et al. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* 2010;17:510–22.
40. Pfeifer GP, Rauch TA. DNA methylation patterns in lung carcinomas. *Semin Cancer Biol* 2009;19:181–7.
41. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 2007;448:553–60.
42. Bailey VJ, Keeley BP, Razavi CR, Griffiths E, Carraway HE, Wang TH. DNA methylation detection using MS-qFRET, a quantum dot-based nanoassay. *Methods* 2010;52:237–41.